LREC 2018 Workshop

8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community

PROCEEDINGS

Edited by

Mayumi Bono, Eleni Efthimiou,
Stavroula-Evita Fotinea, Thomas Hanke,
Julie Hochgesang, Jette Kristoffersen,
Johanna Mesch, Yutaka Osugi

ISBN: 979-10-95546-01-6
EAN: 9791095546016

12 May 2018

Proceedings of the LREC 2018 Workshop
"8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community"

12 May 2018 – Miyazaki, Japan

Edited by Mayumi Bono, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie Hochgesang, Jette Kristoffersen, Johanna Mesch, Yutaka Osugi

http://www.sign-lang.uni-hamburg.de/lrec2018/

Organising Committee

- Mayumi Bono, National Institute of Informatics, Tokyo, Japan
- Eleni Efthimiou, Institute for Language and Speech Processing, Athens, Greece
- Stavroula-Evita Fotinea, Institute for Language and Speech Processing, Athens, Greece
- Thomas Hanke, Institute of German Sign Language, University of Hamburg, Germany
- Julie Hochgesang, Gallaudet University, Washington, USA
- Jette Kristoffersen, Centre for Sign Language, University College Copenhagen, Denmark
- Johanna Mesch, Stockholm University, Sweden
- Yutaka Osugi, Tsukuba University of Technology, Japan

Programme Committee

- Mayumi Bono, National Institute of Informatics, Tokyo, Japan
- Annelies Braffort, LIMSI/CNRS, Orsay, France
- Onno Crasborn, Radboud University, Nijmegen, The Netherlands
- Sarah Ebling, Institute of Computational Linguistics, University of Zurich, Switzerland
- Eleni Efthimiou, Institute for Language and Speech Processing, Athens, Greece
- Michael Filhol, CNRS–LIMSI, Université Paris-Saclay, Orsay, France
- Stavroula-Evita Fotinea, Institute for Language and Speech Processing, Athens, Greece
- Sylvie Gibet, IRISA, University of Bretagne Sud, Vannes, France
- Thomas Hanke, Institute of German Sign Language, University of Hamburg, Germany
- Alexis Héloir, DFKI, Saarbrücken, Germany
- Julie Hochgesang, Gallaudet University, Washington, USA
- Matt Huenerfauth, Rochester Institute of Technology, Rochester, NY, USA
- Trevor Johnston, Macquarie University, Sydney, Australia
- Reiner Konrad, Institute of German Sign Language, University of Hamburg, Germany
- Jette Kristoffersen, Centre for Sign Language, University College Copenhagen, Denmark
- John McDonald, DePaul University, Chicago, USA
- Johanna Mesch, Stockholm University, Sweden
- Carol Neidle, Boston University, Boston, USA
- Yutaka Osugi, Tsukuba University of Technology, Tsukuba JP
- Christian Rathmann, Humboldt-Universität zu Berlin, Germany
- Rosalee Wolfe, DePaul University, Chicago, USA

Preface

This collection of papers stems from the Eighth Workshop on the Representation and Processing of Sign Languages, held in May 2018 as a satellite to the Language Resources and Evaluation Conference in Miyazaki (Japan).
While there has been occasional attention to sign languages at the main LREC conference, the main focus there is on spoken languages in their written and spoken forms.
This series of workshops, however, offers a forum for researchers focussing on sign languages.
While the majority of papers deals with corpus data and technology, this year's hot topic "Involving the Language Community" broadens the scope of the workshop a bit, focussing on the "Return on Investment" the language community certainly deserves.

The contributions composing this volume are presented in alphabetical order by the first author. For the reader's convenience, an author index is provided as well.

We would like to thank all members of the programme committee who helped us reviewing the submissions to the workshop within a very short timeframe!

Finally, we would like to point the reader to the proceedings of the previous workshops that form important resources in a growing field of research. They are all available online from http://www.sign-lang.uni-hamburg.de/lrec/.

The Editors

Table of Contents

Table of Contents

Table of Contents

Author Index

Author Index

Modeling and Predicting the Location of Pauses for the Generation of Animations of American Sign Language

Sedeeq Al-khazraji, Sushant Kafle, Matt Huenerfauth
Rochester Institute of Technology, Golisano College of Computing and Information Sciences
152 Lomb Memorial Drive, Rochester, NY 14623, 1-585-475-2459
{sha6709, sxk5664, matt.huenerfauth}@rit.edu

Abstract
Adding American Sign Language (ASL) animation to websites can improve information access for people who are deaf with low levels of English literacy. Given a script representing the sequence of ASL signs, we must generate an animation, but a challenge is selecting accurate speed and timing for the resulting animation. In this work, we analyzed motion-capture data recorded from human ASL signers to model the realistic timing of ASL movements, with a focus on where to insert prosodic breaks (pauses), based on the sentence syntax and other features. Our methodology includes extracting data from a pre-existing ASL corpus at our lab, selecting suitable features, and building machine learning models to predict where to insert pauses. We evaluated our model using cross-validation and compared various subsets of features. Our model had 80% accuracy at predicting pause locations, out-performing a baseline model on this task.

Keywords: American Sign Language, Animation Synthesis, Prosodic Breaks, Pauses, Modeling

1. Introduction

American Sign Language (ASL) is used as a primary means of communication for about one half million people (Mitchell et al., 2006). ASL is a natural language that consists of movements of the hands, head, body, and face to convey meaning, and it has its a syntax, word order, and lexicon which are distinct from spoken English. While there is great diversity in the English reading literacy skills among members of the Deaf community in the U.S., including some individuals with strong skills, there are also many individuals with lower English reading skills, due to reduced levels of language-exposure during childhood or other educational circumstances. In fact, standardized testing has shown that the majority of Deaf high school graduates in the U.S. (students who are completing secondary school, typically age 18) have English reading skills at the "fourth-grade" level (age 9 students in the U.S.) (Traxler, 2000). Because of the linguistic differences between ASL and English, there are people fluent in ASL but with difficulties reading English text.

Amid these literacy issues, some English text on websites may be too difficult to read. While adding videos of a human signer to websites may sound like a simple solution, this is impractical: Online information is often updated or generated automatically based on a query. A video would need to be recorded and uploaded, which would be costly and time-consuming. Professional animators can produce realistic animations of virtual humans, but the process is also slow. For these reasons, many researchers, e.g. (Adamo-Villani and Wilbur, 2015; Cox et al., 2002; Ebling and Glauert, 2016; Huenerfauth, 2004; Jennings et al, 2010; Kacorri, 2016; Kennaway et al., 2007; Lu, 2014; McDonald et al. 2016; Segout and Braffort, 2009), investigate the development of software that can generate understandable ASL animations of a virtual human signer automatically from an easy-to-update script. The challenge is that this software must configure the animation so that the movements are accurate and easily understood by ASL signers (Huenerfauth, 2008; Huenerfauth and Lu, 2010).

The primary focus of this paper is to investigate using motion-capture data that our lab has previously recorded from human signers to build predictive models for inserting pauses in ASL animations. Prior studies at our lab (Huenerfauth, 2009) have shown that adding linguistically motivated pauses and adjusting the duration of signs enhances the understandability of ASL animation (as measured on a comprehension task). Thus, our goal is to automate this aspect of animation synthesis and to create understandable ASL animation with better quality.

2. Literature Review

2.1 Linguistic Research on Pausing

Some prior psycholinguistic studies have focused on the timing and pausing in ASL and spoken English (Grosjean and Lane, 1977; Grosjean, 1977; Grosjean et al., 1979). For example, Grosjean et al., (1979) investigated the pause length and location of pauses in ASL, based on the sentence structure. Others studied the sign duration and sign speed, based on video observation, and they analyzed speaker and signer performances at different rates (Grosjean, 1979). Grosjean and Lane (1977) found that for spoken English, longer pauses take place at sentence boundaries (pause length longer than 445 ms); shorter pauses take place between noun phrases, verb phrases and conjoined sentences (pause length range between 245 and 445 ms), and the shortest pauses occurred within phrasal constituents (pause length less than 245 ms). For ASL, Grosjean and Lane (1977) analyzed videos to estimate an average pause length: between sentences (229 ms), between conjoined sentences (134 ms), pauses between noun or verb phrases (106 ms), within verb phrases (11 ms), and within noun phrases (6 ms). These findings suggested that pause length in ASL was related to the syntax structure of the sentence, and these findings have inspired the selection of features for our models, as discussed in section 4.2.

2.2 Rule-Based Pausing for Signing Animation

In some prior sign language animation systems, the speed and duration of signing are invariant or must be specified by a human authoring the message: The eSign project developed an animated avatar that performed sequences of signs from a lexicon. To specify the duration of each word, the authors examined the speed of human signers in videos performing each sign (Kennaway et al., 2007), but they did not vary the duration of signs based on where they appeared in sentences. Sign Smith Studio was a commercial product for ASL animation generation; it allowed the user to modify the temporal parameters of the signs in a sentence manually (by adjusting numerical values for the transition time between words, the hold/pause at the end of words, and a multiplier factor for hand speed during the sign). Therefore, a Sign Smith Studio user needed some skill in computer animation and instincts about how to set numerical values for ASL timing (Vcom3D, 2017).

Many projects have implemented sets of rules that govern the speed and timing of sign-language animations, e.g.:

- Huenerfauth (2009) built a model for the duration (length) of signs, the location of pauses, and the length of pauses in ASL. However, his model was based on rules he authored based on some published data in the psycholinguistics literature on ASL (summarized above in 2.1). He designed two algorithms: for calculating sign duration time and for calculating pause location and length. His sign-duration algorithm depended on whether specific signs had previously appeared in a passage and whether they were at the end of clauses, e.g. noun signs located at boundaries (sentence or clause) were lengthened in duration by a set percentage (12% and 8% respectively). For verb signs, subsequent occurrences were shortened in duration by 12%. The values used in these rules were based on averages reported in the linguistics literature, not on any data-driven machine-learning method.
- A more recent study by Villani and Wilbur (2015) also utilized a rule-based approach. Their system predicted how to add prosodic enhancements to ASL animations, including insertion of pauses and phrase-final lengthening of sign duration. To determine Pausing, Villani and Wilbur adopted linguistic values from (Pfau et al., 2005) to insert pauses between and within sentences. Regarding Phrase Final Lengthening, they increased the length of the last sign of a phrase based on a prior study by Wilbur (2009). Their initial user evaluation showed promising results from using this algorithmic approach to add prosodic features.
- Ebling and Glauert (2016) built a system for translating train announcements from German text to Swiss German Sign Language using the JASigning animation platform. The authors wrote a rule to insert a short pause after each item in lists, based on a suggestion from deaf users who viewed their system's animation output; however, they did not provide a general rule for when pauses should be inserted nor what the pause duration should be, in novel contexts.

2.3 Data-Driven Sign Language Research

While many advances in computational linguistics have come from data-driven methods based on machine-learning models, most prior work on sign language has been rule-based, because of the small quantity of training data, e.g. available audio/video recordings that have been linguistically-annotated. As additional signing corpora have recently become available, there has been a recent trend among sign-language researchers of applying data-driven approaches, as discussed in (Huenerfauth, 2014). For instance, various researchers have examined data-driven methods for sign language translation research:

- Bungeroth et al. (2006) created a corpus for German Sign Language and studied machine-translation and facial-expression issues, but not speed or timing.
- Morrissey and Way (2005) investigated example-based machine translation approaches for producing sign language from English text, using a corpus they annotated with manual and non-manual features. They generated word sequences for sign language, not any animation output, which would have required speed or timing information (Morrissey and Way, 2005). Most of these prior studies made use of small corpora containing texts on a special topic/domain, and none of them explicitly modeled speed and pausing of signs.
- Naert et al (2017) investigated automatically adjusting manual segmentation of sign language motion data.

Some researchers have used data-recordings from humans to generate animation output, for example:

- Segouat and Braffort (2009) used rotoscopy to create a French Sign Language corpus and built an animation system that combined different elements of human motion to create novel sentences. While they studied co-articulation (how the movements at the end of one sign are influenced by the beginning of the next), they did not model speed or timing issues directly.
- Cox et al. (2002) built and evaluated a system called "TESSA" for converting English speech to British Sign Language (BSL) animations, using some template-like phrases to build a limited set of sign language sentences. Since their system filled words into templates (rather than synthesizing complete phrases), they did not address timing and pausing issues, which is the focus of our current work.
- In prior work, our lab has used our ASL Motion-Capture Corpus (Lu and Huenerfauth, 2012; 2014) to investigate different aspects of ASL animation: inflecting verb movement (Lu, 2014), facial expression (Kacorri, 2016), and spatial reference point locations (Gohel, 2016).

3. Research Question

While there has been prior translation and animation-synthesis research that has utilized data-driven techniques, as described above, there has not been prior work that has utilized motion-capture corpora of ASL to directly train machine-learning models of speed, timing, or pause-insertion. Given the success of these prior projects (focusing on other aspects of animation) at using motion-capture recordings to build models of how human ASL signers behave, we therefore intend to use a similar method to investigate the following research question:

Research Question: Can we accurately predict where human signers insert pauses in their ASL signing, as evaluated via cross-validation on an annotated corpus of human ASL signing?

Table 1: Detailed information about selected features

Feature name	Explanation	Type
Sentence Boundary (SB), Clause Boundary (CB), Noun Phrase Boundary (NPB), Verb Phrase Boundary (VPB)	Is this inter-sign gap at the boundary of a sentence, clause, noun phrase or a verb phrase?	Categorical: {Yes, No}
Relative Proximity (RP)	How far is this inter-sign gap from midpoint of the current sentence? A detailed formula for calculating this value appears in Huenerfauth (2009).	Numerical
Complexity Index (CI)	The number of syntactic nodes that dominated this inter-sign gap. A detailed formula for calculating this value appears in Huenerfauth (2009).	Numerical
Sentence Length (SL), Noun Phrase Length (NPL), Verb Phrase Length (VPL)	Number of words in the current sentence, the current noun phrase (if applicable), or the current verb phrase (if applicable) at this inter-sign gap position in the corpus.	Numerical

4. Methodology

4.1 Data Preparation

To support our modeling work, we first needed to process and extract relevant information from our existing ASL Motion-Capture Corpus (Lu and Huenerfauth, 2014), which contains: *motion-capture movement data* available as .bvh files (Biovision hierarchical data, XML files representing human joint angles from a movement recording) and *linguistic annotation* (text files exported from the SignStream annotation tool (Neidle, 2002). A team of annotators that included Deaf native ASL signers and linguists, labeled the glosses and syntactic constituents (including sentence, clause, verb phrase, and noun phrase boundaries) in the ASL video and motion-capture corpus, using a process whereby two annotators independently annotated each file and met to discuss their annotations to arrive at a consensus annotation. The annotation included word labels, clause boundaries, and other syntactic information (Lu and Huenerfauth, 2014).

To process this data for our analysis, we wrote Python code to extract timing information and a subset of linguistic annotation properties to produce a comma-separated values (CSV) file with each row representing an inter-word "gap" location, after each of the 7138 words in the corpus, where a prosodic pause could potentially occur. One column was a "target" label that indicated whether this gap location in the corpus was where the human performed a "pause."

Since the original linguistic annotators did not specifically label which inter-sign gaps contained a prosodic pause and which did not, we needed to fill this value automatically: by identifying a threshold time duration to distinguish between regular end-of-sign "holds" (some signs end with the hands remaining in position for a moment) and longer prosodic-break "pauses" during signing. To calculate this threshold, we calculated mean hold time at the end of words, and we subtracted this value from the period of time when the hands were motionless at the end of words. After ranking these durations, we labeled the longest 25% as "pauses" and the remainder as "not pauses," following the typical ASL ratio of prosodic pauses in (Grosjean, 1979).

4.2 Feature Engineering

The remaining columns contain "predictor features," i.e. properties about this gap location (e.g., is this inter-sign gap a boundary between two sentences, what is the length of the current sentence, etc.) that may be relevant to predicting pauses. These features were calculated automatically from annotation present in the original corpus. In summary, our training data set consisted of nine predictor columns and one target column; there were 7138 rows representing gap locations after each ASL sign. Table 1 lists the predictor features implemented in this work and explains their meaning. The various boundary, relative proximity, and complexity index features listed in the table were included based on their use in determining pauses in prior linguistic work (Huenerfauth, 2009; Grosjean et al., 1979). As mentioned in Table 2, detailed formulas for some of these features are described in Huenerfauth (2009), as they were a key part of that prior rule-based model. All of the numerical features were scaled using unity-based normalization with the training minimum and maximum.

The reader may note that none of the features included in our model were lexically specific, i.e. they did not depend on the specific gloss/word labels for the individual signs that preceded or followed any inter-sign gap. This decision to avoid lexically-specific features was intentional, given the relatively small size of our training corpus. Furthermore, a small set of prompts had been used in the collection of this corpus (Huenerfauth and Lu, 2014); thus, we sought to avoid training a model of pause insertion that would be overly domain specific, given our limited data.

4.3 Selecting the Classification Models

Since our goal was to fit and test a model to predict pause locations in ASL animation and our target variable had values of ("there is a pause here" or "there is not a pause here"), we considered a traditional supervised classification approach to make an individualized prediction for the gap following each word in a sentence. Since we had both categorical and numerical predictor features (see the "Type" column in Table 1), we chose to investigate and compare several machine-learning algorithms that support mixed features, including: decision trees, support vector machines (SVM). In particular, we noted that prior work on pause prediction for English (Sarkar and Rao, 2015) or other modeling for ASL (Shibata et al., 2016) had successfully used decision-tree-based learning methods.

To select the optimal subset of features to use when building our model, we implemented code to exhaustively build and test versions of each model using all possible combinations of our predictor features, for a total of 511 different feature subsets. We trained a decision tree classifier (using a maximum of 100 branch nodes) and an SVM Linear classifier in MATLAB (MathWorks, 2017).

Aside from making independent predictions of the target variable ("pause" or "no pause") for each inter-sign gap location, we also investigated if there were dependencies between the values at subsequent gap locations. Specifically, we considered making predictions based on a +/-1 context window (i.e. the predictor features of the inter-sign gap immediately preceding and following the current inter-sign gap), thereby treating the problem as a sequence-tagging problem. For this purpose, we trained a Linear-Chain Conditional Random Field (CRF) model which operated on the context-features and looked for the most optimal path through all possible target sequences for a sequence of words in a sentence.

4.4 Cross-Validation Training and Evaluation

For the classifiers described above, we implemented a 5-fold cross-validation procedure, dividing our data into 80% training set and 20% testing set at each evaluation fold. We calculated the average accuracy and f-score across the 5 folds. To select the best working parameters for each of our models, we performed a grid-search to optimize the model performance. We compared our result with some baselines:

- **Baseline 1:** We inserted a pause at the end of every sentence (and nowhere else). The rationale is that if a human were to create an animation and manually chose to insert some pauses, the animator may likely put them at all of the sentence boundaries, as a simple approach.
- **Baseline 2:** We inserted a pause randomly at 25% of locations. To account for variation due to randomness, we ran it ten times (Table 2 presents the average).

5. Results Analysis

Table 2 shows the accuracy and f-score for each model with best the performing feature combinations. As shown in the table, Baseline 1 (which inserted a pause at all sentence boundaries) has good performance – which is expected as many pauses do occur at the end of sentences.

Table 2: Results of Each Pause-Prediction Classifier

Classifier	Accuracy	F-Score	Features
Linear-Chain CRF	**0.80**	0: 0.298 1: **0.880**	ALL
Decision Tree	0.76	0: 0.226 1: 0.858	CB, VPB
SVM (Linear)	0.76	0: 0.160 1: 0.868	CB, VPB
Baseline 1	0.77	0: **0.392** 1: 0.860	SB
Baseline 2	0.64	0: 0.227 1: 0.768	N/A

The SVM and Decision-tree models, which utilized features from the current inter-sign gap only, struggled to beat this baseline, in both accuracy and f-score. The linear-chain CRF model was our top performing model, with an accuracy of 80% and F-score comparable in performance to (and slightly exceeding) the Baseline 1.

6. Conclusions

In this work, we demonstrated our methodology for building models of one aspect of ASL animation timing, based on machine-learning modeling of a collection of motion-capture data. Specifically, our work has focused on building models of where people pause during signing, and we have successfully identified a set of features and a modeling approach that outperforms a commonly-used baseline for pause-placement (i.e. insert a pause at every sentence boundary). Notably, we have presented a model that utilizes a set of features related to the syntax structure of a sentence (rather than utilizing lexically specific features, such as word labels), which has enabled us to make use of a relatively small corpus to train our model.

We envision that this model could be incorporated as part of a system for automatically synthesizing animations of sign language, with the assumption that such a system is aware of the location of syntactic phrase boundaries during the generation of sentences (which is the basis for all features listed in Table 1), and thereby our model could utilize this information to automatically determine where to insert pauses in the resulting sign-language animation.

In future work, we plan to investigate additional predictive features and modeling techniques for this task, and to conduct a user-based study (with ASL signers evaluating the quality of animations resulting from this model). In subsequent work, we plan to investigate models of the duration (length) of both pauses and individual signs, with an ultimate goal of building software that can generate realistic and understandable animations of ASL, to make information more accessible for ASL signers who may prefer to receive information in the form of sign language or may have reduced reading literacy in written language.

7. Appendix

The Decision Tree and SVM classifiers were implemented in MATLAB using the *Classifier Learner Package*[1], while the Linear-Chain CRF classifier was implemented using the *sklearn-crfsuite*[2] package in Python. Table 3 displays the parameter settings used to build the respective models.

Table 3: Parameters for machine-learning models

Classifier	Function	Parameters
Linear-Chain CRF	*CRF*	algorithm: l2sgd c2: 0.0869 max_iterations: 100 all_possible_transitition: True
Decision Tree	*fitctree*	SplitCriterion: gdi MaxNumSplits: 100 Surrogate: off
SVM (Linear)	*fitcsvm*	KernelFunction: linear PolynomialOrder: [] KernelScale: auto BoxConstraint: 1

[1] https://www.mathworks.com/products/statistics/classification-learner.html

[2] https://sklearn-crfsuite.readthedocs.io/en/latest/

8. Bibliographical References

Adamo-Villani, N., & Wilbur, R. B. (2015, August). ASL-Pro: American sign language animation with prosodic elements. In *International Conference on Universal Access in Human-Computer Interaction* (pp. 307-318). Springer, Cham.

Bungeroth, J., Stein, D., Dreuw, P., Zahedi, M., & Ney, H. (2006, May). A German sign language corpus of the domain weather report. In *Fifth International Conference on Language Resources and Evaluation* (pp. 2000-2003).

Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., & Abbott, S. (2002, July). Tessa, a system to aid communication with deaf people. In *Proceedings of the fifth international ACM conference on Assistive technologies* (pp. 205-212). ACM.

Ebling, S., & Glauert, J. (2016). Building a Swiss German Sign Language avatar with JASigning and evaluating it among the Deaf community. *Universal Access in the Information Society*, *15*(4), 577-587.

Gohel, J. (2016). Modeling the locational distribution of spatial reference points established by ASL signers in a motion capture data corpus. *Master's capston report, R.I.T.*

Grosjean, F. (1977). The perception of rate in spoken and sign languages. *Attention, Perception, & Psychophysics*, *22*(4), 408-413.

Grosjean, F. (1979). A study of timing in a manual and a spoken language: American Sign Language and English. *Journal of Psycholinguistic Research*, *8*(4), 379-405.

Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive psychology*, *11*(1), 58-81.

Grosjean, F., & Lane, H. (1977). Pauses and syntax in American sign language. *Cognition*, *5*(2), 101-117.

Huenerfauth, M. (2004). Spatial and planning models of ASL classifier predicates for machine translation. In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation* (TMI 2004). Baltimore, MD, USA.

Huenerfauth, M. (2008). Generating American Sign Language Animation: Overcoming Misconceptions and Technical Challenges. *Universal Access in the Information Society*, 6(4). Springer-Verlag. DOI=https://doi.org/10.1007/s10209-007-0095-7

Huenerfauth, M. (2009). A linguistically motivated model for speed and pausing in animations of American Sign Language. *ACM Transactions on Accessible Computing (TACCESS)*, *2*(2), article 9, 31 pages. DOI=http://dx.doi.org/10.1145/1530064.1530067

Huenerfauth, M. (2014). Learning to generate understandable animations of American sign language. In *Proceedings of the 2nd Annual Effective Access Technologies Conference*, Rochester, NY, USA, June 2014. *R.I.T.*

Huenerfauth, M., & Kacorri, H. (2015). Augmenting EMBR virtual human animation system with mpeg-4 controls for producing ASL facial expressions. In *International Symposium on Sign Language Translation and Avatar Technology* (3).

Huenerfauth, M., & Lu, P. (2010). Modeling and synthesizing spatially inflected verbs for American Sign Language animations. In *Proceedings of the 12th international ACM SIGACCESS conference on computers and accessibility (ASSETS'10), 99-106. DOI=http://dx.doi.org/10.1145/1878803.1878823*

Jennings, V., Elliott, R., Kennaway, R., & Glauert, J. (2010, May). Requirements for a signing avatar. In *Proceedings of the 4th LREC Workshop on the Representation and Processing of Sign Languages* (pp. 133-136).

Kacorri, H. (2016). *Data-driven synthesis and evaluation of syntactic facial expressions in American Sign Language animation*. City University of New York.

Kennaway, J. R., Glauert, J. R., & Zwitserlood, I. (2007). Providing signed content on the Internet by synthesized animation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *14*(3), 15.

Lu, P. (2014). *Data-driven synthesis of animations of spatially inflected American Sign Language verbs using human data*. City University of New York.

Lu, P., & Huenerfauth, M. (2012). CUNY American Sign Language motion-capture corpus: first release. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, The 8th International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey*.

Lu, P., & Huenerfauth, M. (2014). Collecting and evaluating the CUNY ASL corpus for research on American Sign Language animation. *Computer Speech & Language*, *28*(3), (May 2014), pp. 812-831. Elsevier. DOI=https://doi.org/10.1016/j.csl.2013.10.004

MathWorks Inc. 2017. Classification Learner-MATLAB. MATLAB and Statistics Toolbox Release 2017a. Retrieved April 13, 2017 from https://www.mathworks.com/products/statistics/classification-learner.html

McDonald, J., Wolfe, R., Schnepp, J., Hochgesang, J., Jamrozik, D. G., Stumbo, M., Berke, L., Bialek, M. & Thomas, F. (2016). An automated technique for real-time production of lifelike animations of American Sign Language. Universal Access in the Information Society 14(4):2016. pp. 551-566.

Mitchell, R. E., Young, T. A., Bachleda, B., & Karchmer, M. A. (2006). How many people use ASL in the United States? Why estimates need updating. *Sign Language Studies*, *6*(3), 306-335.

Morrissey, S., & Way, A. (2005). An example-based approach to translating sign language. *Proc. Workshop on Example-Based Machine Translation,* 109-116.

Naert, L., Larboulette, C., & Gibet S. (2017). Coarticulation Analysis for Sign Language Synthesis.

International Conference on Universal Access in Human-Computer Interaction, Vancouver, Canada.

Neidle, C. (2002). SignStream™: A Database Tool for Research on Visual-Gestural Language. In Brita Bergman, Penny Boyes-Braem, Thomas Hanke, and Elena Pizzuto, eds., Sign Transcription and Database Storage of Sign Information, a special issue of Sign Language and Linguistics 4 (2001):1/2, pp. 203-214.

Pfau, R., Steinbach, M., & Woll, B. (Eds.). (2012). *Sign language: An international handbook* (Vol. 37). Walter de Gruyter.

Sarkar, P., & Rao, K. S. (2015, July). Data-driven pause prediction for synthesis of storytelling style speech based on discourse modes. In Electronics, Computing and Communication Technologies (CONECCT), 2015 IEEE International Conference on (pp. 1-5). IEEE.

Segouat, J., & Braffort, A. (2009, February). Toward the study of sign language coarticulation: methodology proposal. In *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on* (pp. 369-374). IEEE.

Shibata, H., Nishimura, H., & Tanaka, H. (2016, July). Basic Investigation for Improvement of Sign Language Recognition Using Classification Scheme. In International Conference on Human Interface and the Management of Information (pp. 563-574). Springer International Publishing.

Traxler, C. B. (2000). The Stanford Achievement Test: National norming and performance standards for deaf and hard-of-hearing students. *Journal of deaf studies and deaf education*, *5*(4), 337-348.

Vcom3D, Inc, Sign Smith. Retrieved December 13, 2017 from http://www.vcom3d.com/signsmith.php

Wilbur, R. B. (2009). Effects of varying rate of signing on ASL manual signs and nonmanual markers. Language and speech, 52(2-3), 245-285.

Tactile Japanese Sign Language and Finger Braille:
An Example of Data Collection for Minority Languages in Japan

Mayumi Bono[1&2], Rui Sakaida[1], Ryosaku Makino[3], Tomohiro Okada[2], Kouhei Kikuchi[2], Mio Cibulka[4], Louisa Willoughby[5], Shimako Iwasaki[5], and Satoshi Fukushima[6]

[1] National Institute of Informatics, [2] SOKENDAI (The Graduate University of Advanced Studies), [3] Waseda University, [4] Lund University, [5] Monash University, [6] The University of Tokyo

[1] 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 JAPAN, [2] Shonan Village, Hayama, Kanagawa 240-0193 JAPAN, [3] 2-579-15 Mikajima, Tokorozawa city, Saitama 359-1192 JAPAN, [4] Helgonabacken 12, SE-223 62 Lund SWEDEN, [5] Victoria 3800 AUSTRALIA, [6] 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904 JAPAN

{bono, lui, tokada-deaf}@nii.ac.jp, rmakino@aoni.waseda.jp, kikuchi_kouhei@soken.ac.jp, Cibulka@posteo.de, {louisa.willoughby, shimako.iwasaki}@monash.edu, fukusima@rcast.u-tokyo.ac.jp

Abstract

We recorded data on deafblind people in Japan. In this filming project, we found that Japanese deafblind people use different communication methods, tactile Japanese sign language and finger braille, depending on their hearing ability and eyesight. Tactile sign language is normally used by those who were born deaf or lost their hearing at an early age and then lost their sight after acquiring a sign language. These people are known as deaf-based deafblind (D-deafblind). Finger braille is popular in Japan, but largely unknown elsewhere. It is normally used by those who were born blind or lost their sight at an early age and subsequently lost their hearing after learning how to produce speech using their throat and mouth. These people are known as blind-based deafblind (B-deafblind hereafter). This paper introduces our filming project; the ways of data collection, translation and annotation. In addition, we show our preliminary observations using our data sets to clarify the important fact that we should collect their interactions at this moment. The data show how their interactions have already become established and sophisticated in their communities. We discuss how our filming project will contribute to the deafblind community in Japan.

Keywords: Tactile Japanese Sign Language (Tactile JSL), Finger Braille, minority language, data collection

1. Introduction

This paper introduces our data-collection project and preliminary observations on two minority languages used in Japan: tactile Japanese sign language and finger braille. Tactile sign language is normally used by those who were born deaf or lost their hearing at an early age and then lost their sight after acquiring a sign language. These people are known as deaf-based deafblind (D-deafblind hereafter). Finger braille is popular in Japan, but largely unknown elsewhere. It is normally used by those who were born blind or lost their sight at an early age and subsequently lost their hearing after learning how to produce speech using their throat and mouth. We call these people blind-based deafblind (B-deafblind hereafter).

2. Background

Based on a survey conducted every 5 years by the Ministry Health, Labor, and Welfare, there are an estimated 23,000 deafblind people in Japan. The Tokyo Deafblind Club (Tokyo *mourousha tomo no kai*, a certified non-profit organization)[1], which provides social services such as dispatching interpreters and volunteers, reported that there were 840 deafblind people with disability certificates in Tokyo in 2012. However, only 10% of them were registered as members of the club.

2.1 Tactile Japanese Sign Language (Tactile JSL)

Tactile sign language enables deafblind people to communicate through touching. To communicate with others after losing their sight, most D-deafblind people convert from visual sign language to tactile sign language. In one standard method, the deafblind signer's hands are positioned under the recipient's hands when signing (Fig. 1).

Mesch (2001) suggested that deafblind users of tactile Swedish sign language use different hand configurations for tactile reception depending on whether the interaction is primarily monologic or dialogic. In the dialogic position, the signer receives signs with his/her right hand under the interlocutor's left hand, while the interlocutor receives signs with his/her right hand under the signer's left hand. In this crossed situation, it is possible for both to be signers and to take turns, and to speak smoothly. In Japan, however, deafblind people rarely use the dialogic position to converse with other deafblind people. Of course, it is possible for very experienced tactile JSL signers and people who are born deafblind to use the left hand to receive information from an interpreter's right hand.

We need to collect tactile JSL data, including situations involving D-deafblind people, to compare their diverse

Figure 1: An image of Tactile JSL dialogue

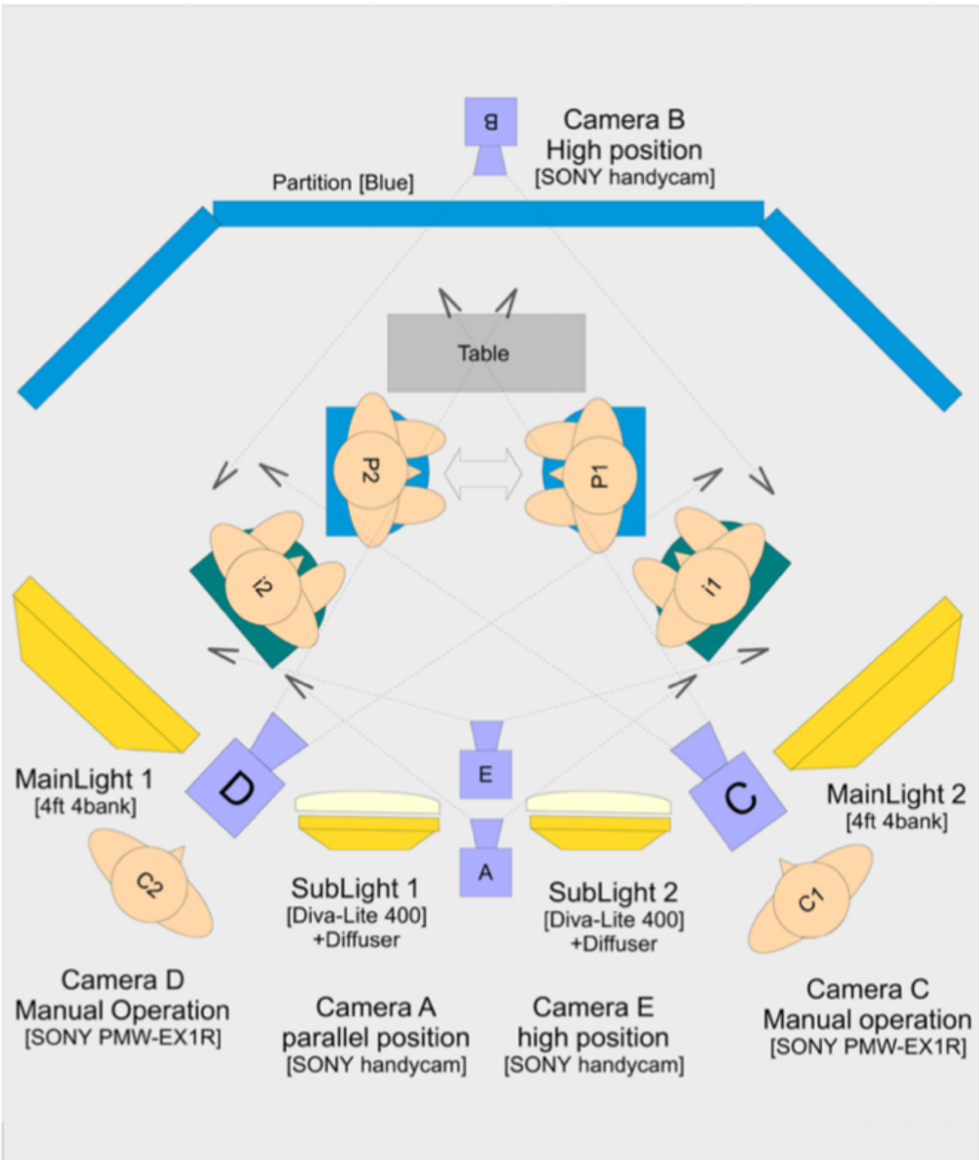
[1] http://www.tokyo-db.or.jp/

Figure 2: An image of a Finger Braille interaction via interpretation

methods of communication with those used in other countries.

2.2 Finger Braille

Finger braille was developed by Prof. Fukushima (B-deafblind) and his mother. Finger braille users and interpreters tap the index, middle, and ring fingers of both hands of the B-deafblind person, like tapping on a braille typewriter in a mora-by-mora manner.
The interpreter sits beside the deafblind person so that he/she can place his/her hands in the same direction as the hands of the deafblind person (Fig. 2). Prof. Fukushima reported that highly skilled interpreters not only tap the fingers to indicate letters but also represent the speaker's stance and attitude with the tapping pressure and rate.



<Camera data>
Size: HQ1280/59.94p
W: 3200K
F: T2 / F1.9-2.4
Gain: 0db
SHT: 1/100

Participant : P1, P2
Interpreter : i1, i2
Cameraman : C1, C2

Figure 3: The positions of the cameras and lights

No study has focused on tactile finger interactions from the perspectives of linguistics and communication. We need to collect finger braille interaction data, including interpreting situations, to determine how individuals convey information in a dialogue through finger tapping.

3. Data

In this section, we introduce the filming procedure, dataset, camera settings, and the number of subjects.
Our study was approved by the ethics committee of the National Institute of Informatics.

3.1 Filming: Tactile JSL Interaction

To collect tactile JSL interactions, we filmed individuals in (1) experimental and (2) natural settings.

3.1.1 Experimental Setting

For the experimental setting, we built a temporary studio at the National Museum of Emerging Science and Innovation (Miraikan), Odaiba, Tokyo (Figs. 3). We used four cameras: two professional cameras (SONY-PMW-EX1R) on tripods that were operated manually to detect the hand movements and two SONY Handycams on tripods set at high angles to detect spatial orientation and locational information, such as sitting position and distance between individuals. We simultaneously recorded the interpreters' spoken words to make it easier for the annotators to find and tag interesting parts in the dialogues. We used four professional lighting devices: two main lights (4 feet, bank of four) and two sub-lights (Diva-Lite 400 and diffuser).
We filmed the deafblind dialogues of two pairs of subjects: pair A was a male–female pair and pair B was a male–male pair. All of the individuals were familiar with each other. They had been introduced to our project by the Tokyo Deafblind Club. The total length of the recordings was 1 hours and 21 minutes. These recordings are suitable for observing the organization of the turn-taking system (Sacks *et al*., 1974) in tactile JSL.

3.1.2 Naturally-occurring Conversations

The Tokyo Deafblind Club hosts several events to encourage social relationships among deafblind people, interpreters, and volunteers. These include sightseeing, gardening lessons, and gymnastics. In July 2015, the club visited the National Diet Building in Tokyo. Because we could not get permission to film inside the building, we filmed their interactions after the sightseeing tour outside the Diet using two SONY Handycams.
The group included 10 deafblind people, 20 interpreters, and some of the club staff. We filmed various conversations, mainly those including deafblind people in conversation with deafblind, deaf, narrow-viewed deaf, and hearing individuals via an interpreter (Fig. 4). When we noticed a deafblind person starting to chat with somebody, we started filming all of the participants from one angle. The total length of the recordings was approximately 17 minutes. These recordings are suitable for observing the organization of the F-formation system (Kendon, 1990) in tactile JSL, tactile JSL via an interpreter, and tactile JSL leaners.

Figure 4: An image of data collection outside the National Diet Building, Tokyo.

3.2 Filming: Finger Braille Interaction

We filmed tactile JSL interactions in (1) interviews and (2) natural settings.

3.2.1 Interview Settings

We had the opportunity to film finger braille dialogue via interpretation in January 2015 when we were writing an article for the Information-Processing Society of Japan. The dialogue was an interview in which the first author of this paper interviewed Prof. Satoshi Fukushima, who is B-deafblind and developed finger braille with his mother and the finger braille interpretation method with his friends. He positioned two highly skilled interpreters on either side of him, while the interviewer sat in front of him (Fig. 2). The interview lasted almost 3 hours and covered many topics, focusing on his daily communication with others, email, chatting, and dealing with the Japanese government to ask for support for the deafblind community. The recording was 2 hours and 18 minutes long and is suitable for observing the organization of the turn-taking system (Sacks *et al*., 1974) in finger braille interaction via interpreters.

3.2.2 Naturally-occurring Conversations

The Tokyo Deafblind Club held a communication-exchange event in Tokyo in November 2017 (Fig. 5). The purpose of this event was to exchange methods of communication, such as tactile JSL, finger braille, hand writing, and small-size signing via a narrow view, to encourage peoples' understanding of each other. Ten deafblind people, 20 interpreters, four guests (our project members), and some of the club staff attended this event.
The deafblind people with interpreters were divided into two groups: five waiting groups and five visiting groups. Five tables were placed in a circle in a room. The members of the waiting groups sat on the outer sides of the tables, while the members of the visiting groups sat on the inner sides of the tables. The groups then chatted for 10 minutes. After each chat, the visiting groups moved to the next table. Including a 10-minute break, the event took 60 minutes.

Figure 5: An image of data collection during a communication-exchange event organized by the Tokyo Deafblind Club.

We filmed their interactions using four SONY Handycams operated manually with a monopod. The recordings totaled 6 hours and 17 minutes. These recordings are suitable for observing the organization of the turn-taking system (Sacks *et al*., 1974), including learners of both tactile JSL and finger braille, and via interpreters.

4. Translation and Annotation

We have started to translate the data into Japanese on ELAN,[2] except for the films made in November 2017.

4.1 Tactile JSL

4.1.1 Three Kinds of Translation by Interpreters

We asked some highly skilled, very experienced sign language interpreters (hearing) to translate the data into Japanese idiomatic translation, English idiomatic translation and Japanese word-order translation on text format (Fig. 6, upper left). The idiomatic translation (IT) serves as ideal forms of sentences as language. The word-order translation (WOT) serves to maintain the original word order of sign language. At this level, the text in translation is very consciously written in a grammatically inaccurate manner of Japanese. By keeping this kind of translation, we could always show how much signed language differs from spoken language.

4.1.2 Basic Annotation on ELAN by Deaf Signers

Then, experienced Deaf interpreters put the translations and the information of hand positions on ELAN (Fig.6, right). In this phase, they checked the meanings of the signing, which were put by interpreters (hearing), and put the correct timing of the signing on ELAN format. They also corrected some mistranslations by interpreters.
Furthermore, they annotated the current signer's hand position such as 'both inside (both hand inside), both outside (both hand outside) to frequency of hand position they use in interaction. In Tactile JSL, the dialogue position (Mesch, 2001) is not so often used. If we found them, we put, rh-outside (right hand outside) and so on. At this moment, we have found only one case of rh-inside in our data set.

4.1.3 Focused Annotation: Hand Movement

For the cases for which we had already identified some focal points for analysis, we added specific annotations to the excerpts, such as Gesture Units (McNeill, 1992; Kita *et al*., 1998), gaze direction, and hand position, using Conversation Analysis (CA) style transcription. One of our original points was to establish a physical and hand movement unit smaller than Gloss, called a Movement Unit. We applied the concept of the gesture unit (GU) proposed by Kendon (1972, 1980) to annotate the beginning and end points of signed turns. The GU is the interval between successive rests of the limbs, rest positions, or home positions. A GU consists of one or several gesture phrases. A gesture phrase is what we intuitively call a "gesture," and it, in turn, consists of up to five phases: preparation (optional), stroke (obligatory in the sense that a gesture is not said to occur in the absence of a stroke), retraction (optional), and pre- and post-stroke hold phases (optional).
When analyzing overlapping communications in conversations, it is important to note the timing of the expressions of both the signer and recipient. In signed conversations, articulation involves hand signs that appear in front of the participants; this process of articulation is comparable to the visible lip movements made by those involved in spoken conversations. Using this methodology, we can observe how participants engage in an articulation phase in which signers move their hands to the signing space from the home position as a signal for the start of turn-taking in interactions.

4.2 Finger Braille

4.2.1 Basic Annotation on ELAN

By using interview data mentioned in 3.2.1, we asked a highly skilled and very experienced annotator for spoken language to annotate temporal relationship between interviewer's speech (Bono, the first author) and interviewee's speech (Prof. Satoshi Fukushima, B-deafblind) on ELAN in Japanese, Roman-alphabet style of Japanese, and English idiomatic translation (Fig. 7).

4.2.2 Focused Annotation: Interpretation and Laughing

Then, we annotated the three conditions of interpreters' finger movements such as, *holding, interpreting* and *conveying laughter/nod* (Fig. 7). Holding means the condition in which interpreters' fingers are maintained over B-deafblind person's fingers during silence in dialogue or B-deafblind person's speech. Interpreting means the condition in which interpreters' fingers are tapping on B-deafblind person's fingers to convey interlocutor's speech in a mora-by-mora fashion. Conveying laughter/nod means the condition in which interpreters' fingers are moving in a specific way to represent the current speaker is laughing or the speech including voice quality of laughing.

[2] https://tla.mpi.nl/tools/tla-tools/elan/

01> Both Inside
(WOT-jp) 遊びに行く、行く、行きました。私の趣味なんだけど。
(IT-jp) いろいろなところに行きましたよ。趣味なんです。
(IT-en) I went to a lot of places. It's my hobby.

02> Both Outside
(WOT-jp) あなたは日本と海外のどちらですか？
(IT-jp) 国内と海外、どちらに行かれますか？
(IT-en) Inside Japan or overseas, where did you go?

01> Both Inside
(WOT-jp) 日本、日本だけ。
(IT-jp) もっぱら国内です。
(IT-en) Only inside Japan.

02> Both Outside
(WOT-jp) （とんとん）
(IT-jp) そうですか。
(IT-en) (Tapping) I see.

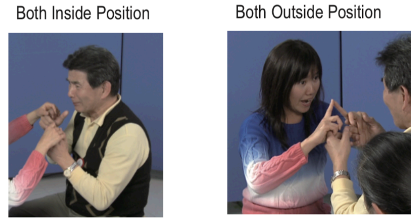


Figure 6: Three kinds of translation (left) and basic annotation of tactile JSL dialogue on ELAN (right).

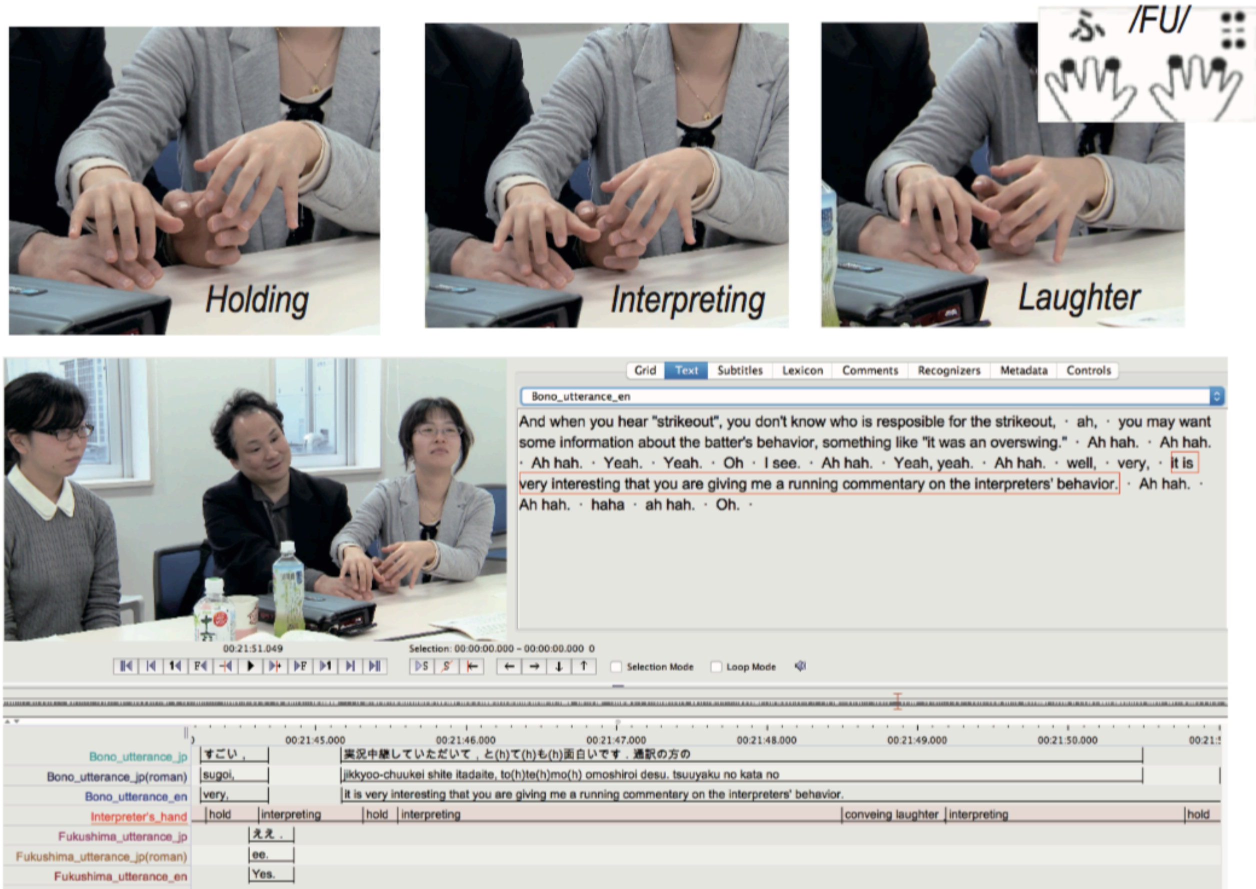


Figure 7: Basic annotation and focused annotation of finger braille interaction on ELAN.

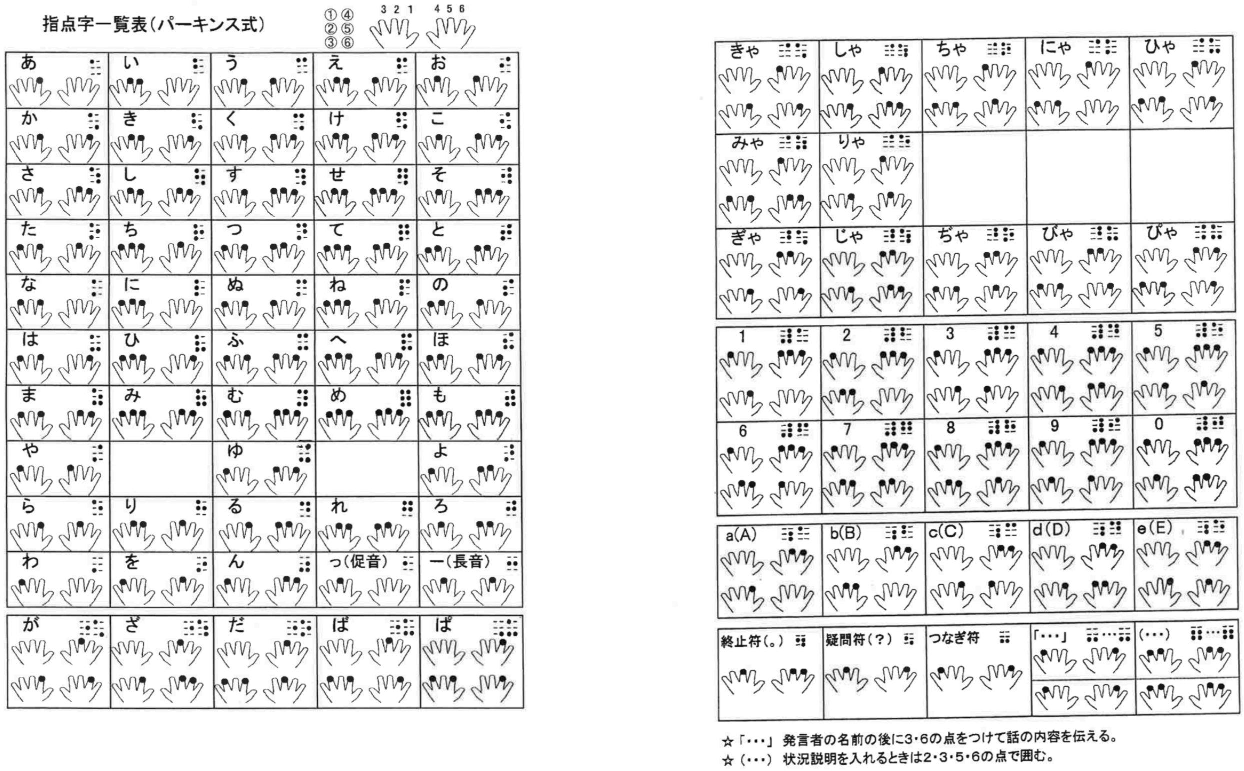


Figure 8: Table of tapping combination of finger braille provided by Ms. Setsuko Sugiura and Mr.Yuji Tako (The Tokyo Deafblind Club).

4.2.3 Focused Annotation: Tapping in Mora-by-mora Fashion

Currently, we are trying to annotate interpreters' finger tapping movements in a mora-by-mora fashion using the table of tapping combination of finger braille provided by Ms. Setsuko Sugiura and Mr.Yuji Tako (The Tokyo Deafblind Club) (Fig.8)[3]. However, standard high-vision cameras have a limited ability to capture these rapid and complicated movements by fingers. If we could annotate these micro-movements, we could observe how finger braille users caught the signals, such as a part of final particle and interactive features by which participants could anticipate the turn-ending comes, to take next turn in dialogue.

5. Examples of Analysis

5.1 Tactile Japanese Sign Language

First, we focused on tapping in tactile JSL interactions. Mesch (2001) examined tapping feedback in tactile sign language interactions. Based on these observations, we tried to identify some of the original features of tactile JSL.

Visual JSL signers have a lexical form of feedback that uses finger tapping movements repeated twice to represent a clear response, such as 'I know' or 'I agree' (Fig. 9, left) (Bono *et al.*, 2014).

As shown in Fig. 9 (middle), a deafblind signer with 2 years of experience as a tactile JSL signer taps the interlocutor's hands using the index fingers and thumbs of both hands twice, and again in the recipient position.

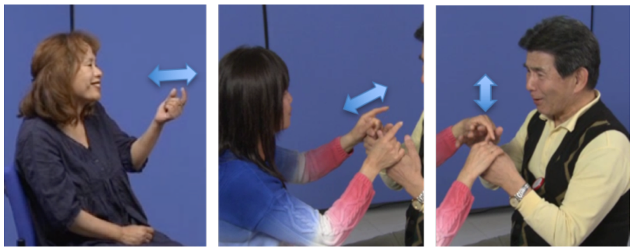


Figure 9: Comparison of finger tapping feedback between visual JSL and tactile JSL

As shown in Fig. 9 (right), a deafblind signer with 16 years of experience as a tactile JSL signer taps the interlocutor's hand using three fingers of the right hand twice, in the speaker position. We are currently analyzing these kinds of tapping feedback to observe the process by which an individual transforms from visual JSL to tactile JSL.

5.2 Finger Braille

In this subsection, observing our experimental-setting data of finger braille interaction between Prof. Fukushima and Bono, we describe how finger braille interpreters convey interlocutor's nonverbal behaviors to the deafblind. As for vocal or signed interpretation, interpreters are not necessarily expected to convey interlocutor's nonverbal behaviors, since the recipient have access to them vocally and/or visually. However, regarding finger braille interpretation, it is crucial for the interpreters to convey interlocutor's nonverbal behaviors when relevant.

3 http://www.tokyo-db.or.jp/wp-content/uploads/2012/10/03679a8282b7c97a65c07ca3ed8620da.pdf

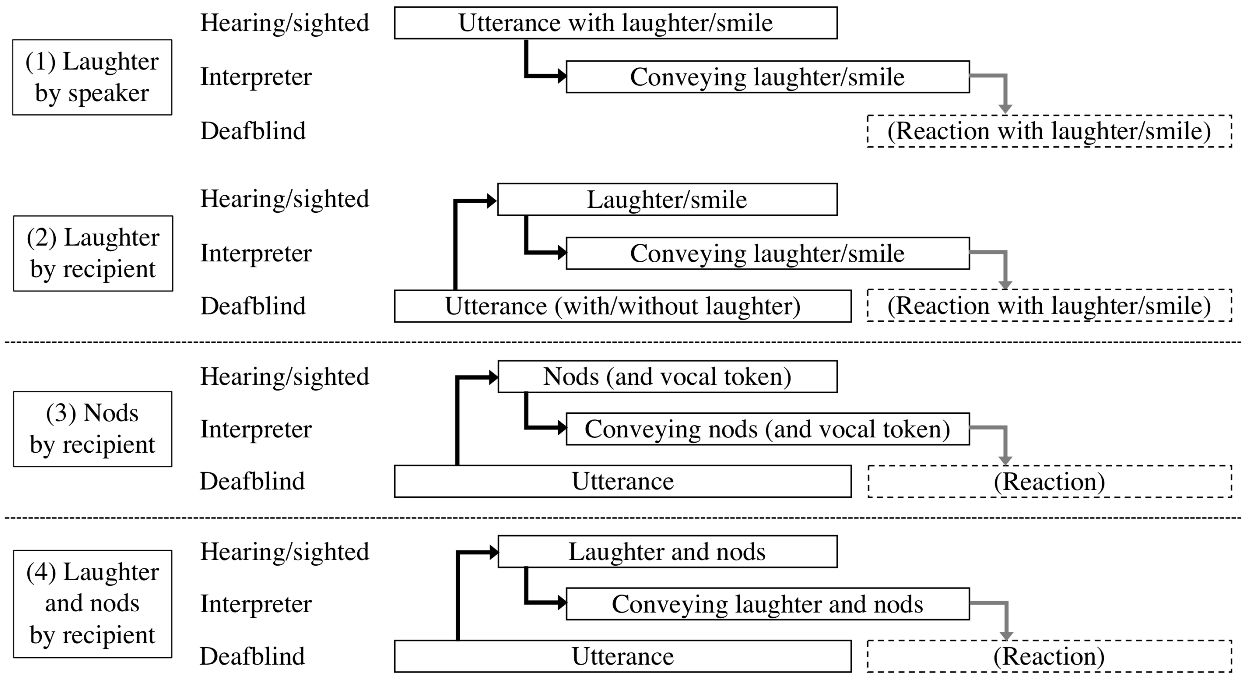


Figure 10: The ways laughter and nods are conveyed in finger braille interpretation.

In particular, finger braille interpreters often convey laughter or nods produced by the hearing/sighted, simultaneously tapping several of the recipients' fingers, several times for laughter, and once or twice for nods. We preliminarily collected and observed several examples in which the interpreters conveyed Bono's nods and/or laughter to Prof. Fukushima. Our tentative observations (Fig. 10) are as follows: Firstly, as for laughter, Bono's laughs within her utterance were frequently conveyed, and sometimes Prof. Fukushima laughed too responding to her (Fig. 10, (1)). In addition, Bono's laughs or smile with slight breath, responding to Fukushima's utterance, were also conveyed (Fig. 10, (2)). Secondly, concerning nods, even when Bono nodded without any vocal utterances, interpreters conveyed her nods to Prof. Fukushima (Fig. 10, (3)). More interestingly, when Bono nodded and slightly later produced a vocal token *un* ("Yeah."), the interpreter conveyed the nods as soon as they started to be produced, that is, before the beginning of *un*. Lastly, when Bono produced both nods and laughter, the interpreter conveyed both of them (Fig. 10, (4)).

Even when the deafblind is a speaker, finger braille interpreters continues to pay attention to the behaviors of the hearing/sighted hearer, and they convey his/her responses, not only their vocal backchannels but also silently produced nods or smiles. On the other hand, when the deafblind is a "hearer", interpreters conveys laughter which the speaker produces, as well as his/her utterances in a mora-by-mora fashion. Under the specific circumstance, finger braille interpreters must be embodied messengers of the speaker, conveying not only linguistically uttered but also bodily expressed messages (cf. Goffman, 1963).

6. Discussion

We found that interaction methods have already become highly established and sophisticated in the deafblind communities.

We have collected several kinds of data on deafblind people in Japan. Through this filming project, we have noticed that the Japanese deafblind people we have observed use different methods of communication, depending on the users' hearing ability and eyesight. As Prof. Satoshi Fukushima, one of the board members of the Tokyo Deafblind Club, insists, we should provide diverse and attentive services, such as training for specialists in tactile JSL and finger braille, that are suitable for the individual needs of deafblind people in their daily communication. There is a tendency for Japanese deafblind people and their communities to respect the communication methods of others, rather than simply standardizing and simplifying their own communication methods.

However, the Japanese deafblind community has not opened up to the World sufficiently. Because it is difficult for Western countries to access written Japanese homepages and documents, deafblind people and their supporters cannot present their activities and thoughts to the World. This situation results in Japanese-specific conditions within the deafblind community, which includes the aforementioned difficulty in communicating their everyday lives to other countries.

In fact, there are some interesting social movements related to tactile sign language in other countries, such as the Pro-tactile movement in Seattle reported by Edwards (2014). It is also difficult for the Japanese deafblind community to access these world-wide social movements. We strongly believe that our filming project provides a good opportunity for them to share their everyday lives with the spoken language community and to connect with other communities in other countries. In particular, finger braille interaction, which was created in Japan, will be made available to B-deafblind people in other countries, bringing with it a strong impact on communication.

7. Conclusions and Future work

This paper introduced our project of filming a deafblind community in Japan. Currently, we are planning to analyze our data from the viewpoints of comparative linguistics studies of visual JSL and tactile JSL, and of spoken Japanese and finger braille.

We have noticed that standard high-vision cameras have a limited ability to capture hand movements in the speaker's position in tactile JSL interaction, where the speaker's hand is located under the recipient's hand. We have already conducted several trials of filming tactile JSL using 360-degree cameras. We hope to collaborate with visual processing engineers when translating and annotating the data.

8. Acknowledgments

This research project was partially supported by KAKENHI (No. 17KT0065), 2017-2019, 'Multimodal Research on Sign Language, Tactile Sign Language and Finger Braille in Interaction: A Bilingual and Bimodal Study of the Influence of Spoken Language', PI: Mayumi Bono, Grant-in-Aid for Scientific Research (B), Generative Research Fields: Orality and Society. The professional cameras and lighting devices were operated by Rakuda Studio.[4]

9. References

Bono, M, Kikuchi, K, Cibulka, P, and Osugi, Y. (2014) Colloquial corpus of Japanese Sign Language: A design of language resources for observing sign language conversations. *Proc. of the 9th Language Resources and Evaluation Conference*, pp. 1898–1904.

Edwards, T. (2014) From compensation to integration: Effects of the pro-tactile movement on the sublexical structure of Tactile American Sign Language, *Journal of Pragmatics*, 69, 22-41.

Goffman, E. (1963) Behavior in Public Places: Notes on the Organization of Gatherings. New York: Free Press.

Kendon, A. (1972) Some relationships between body motion and speech. In Siegman, Aron W. & Pope, Benjamin (Eds.), *Studies in dyadic communication*. 177-210. New York: Pergamon Press.

Kendon, A. (1980) Gesticulation and speech: Two aspects of the process of utterance. In Key, Mary R. (Ed.), *The relation between verbal and nonverbal communication.* The Hague: Mouton.

Kendon, A. (1990) *Conducting Interaction: Patterns of Behavior in Focused Encounters,* Cambridge University Press.

Kita S, van Gijn I., van der Hulst H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. *Lecture Notes in Computer Science*, 1371/1998. Springer Berlin / Heidelberg.

McNeill, D. (1992). *Hand and Mind.* University of Chicago Press, Chicago.

Mesch, J. 2001. *Tactile Sign Language: Turn taking and question in signed conversations of deaf-blind people.* Hamburg: Signum.

Sacks H., Schegloff E., Jefferson G. (1974) A simplest systematics for the organization of turn-taking in conversation. *Language* 50 696–735.

[4] http://www.rakudastudio.com/

Augmenting Sparse Corpora for Enhanced Sign Language Recognition and Generation

Heike Brock[1], **Juliette Rengot**[2], **Kazuhiro Nakadai**[1]
[1]Honda Research Institute Japan Co., Ltd.,[2]Ecole Nationale des Ponts et Chaussées
8-1 Honcho, Wako-shi, Saitama 351-0188, Japan, 6 et 8 avenue Blaise Pascal, 77455 Champs-sur-Marne, France
{h.brock, j.rengot, nakadai}@jp.honda-ri.com

Abstract

The collection of signed utterances for recognition and generation of Sign Language (SL) is a costly and labor-intensive task. As a result, SL corpora are usually considerably smaller than their spoken language or image data counterparts. This is problematic, since the accuracy and applicability of a neural network depends largely on the quality and amount of its underlying training data. Common data augmentation strategies to increase the number of available training data are usually not applicable to the spatially and temporally constrained motion sequences of a SL corpus. In this paper, we therefore discuss possible data manipulation methods on the base of a collection of motion-captured SL sentence expressions. Evaluation of differently trained network architectures shows a significant reduction of overfitting by inclusion of the augmented data. Simultaneously, the accuracy of both sign recognition and generation was improved, indicating that the proposed data augmentation methods are beneficial for constrained and sparse data sets.

Keywords: sign language, machine learning corpus, data augmentation, motion capture data, avatar animation

1. Introduction

In contrast to resources on spoken language, signed machine learning data can hardly be obtained from scratch. Instead, signed conversations need to be recorded first using movement sensing devices such as video or depth cameras, and their content be annotated with the help of experienced or fluent signers. For this reason, corpora for Sign Language (SL) research are seldom as extensive as their spoken counterparts. Corpora for specific tasks such as sign recognition or avatar animation require even more specialized capture environments and settings, consuming a high amount of both financial and human resources. This leads to small and sparse SL corpora that are seldom able to depict all linguistic features under a natural vocabulary scope: common corpora either focus on specific aspects (e.g. facial expressions for sign avatar synthesis (Gibet et al., 2016; Ebling and Huenerfauth, 2015)) or specific language domains (e.g. weather reports (Koller et al., 2015; Umeda et al., 2016)) and usually contain a very small subset of lexical items or separated signs only (Pigou et al., 2014; Ong and Ranganath, 2005).

Especially for the application of modern Machine Translation (MT) methods that base on pure data-driven artificial networks, common SL corpora appear insufficient to learn meaningful data representations. To develop sign recognition and synthesis systems that are robust towards outliers and universally applicable without prior domain knowledge, future data sets need to considerably increase their amount of training data. Here, it appears reasonable to modify existing data in such a way that new data is created that contains the significant characteristics of the signed expressions and their variant lexical items, while simultaneously maintaining their linguistic shape and semantic meaning. However, this requirement excludes the use of common simple data augmentation strategies such as mirroring and flipping or cropping, and new strategies for data augmentation have to be found to develop highly accurate communication technologies in the future.

A number of data augmentation strategies that could be used to increase the number of available training data without data loss are given in this work. Based on a corpus of motion captured sentence expressions in Japanese Sign Language (JSL), we evaluate the benefit of the presented data augmentation methods on the learning of a deep recurrent neural network architecture. Results indicate that problems characteristic for small and sparse data sets can be reduced, suggesting the benefit of the presented strategies for future SL translation interfaces.

2. Machine Learning Corpus Augmentation

Data augmentation is a common machine learning strategy to improve accuracy of classifiers and predictors. Its deployment bases on the presumption that machine learning models do not generalize well when they are trained on data sets that do not contain sufficient variation within the data. Previous research utilizing image training data has shown that data augmentation can act as a regularizer and prevent overfitting of the neural networks (Simard et al., 2003; Cireşan et al., 2010): the more data a machine learning model has access to, the more useful information can be extracted from the original data set and hence the more effective it can be. In line with these results, the most accurate image classification or segmentation networks presented within the last years all utilize techniques to artificially synthesize new samples from existing ones, such as the Imagenet (Krizhevsky et al., 2012) or its succeeding, even deeper network variations (Szegedy et al., 2015; He et al., 2016). Consequently, data augmentation should be particularly useful when the number of available training data is small and new training data cannot be acquired easily, as it is common for sign language corpora. This holds especially true when deep neural networks – that rely largely on the number of available sample data for network learning – are applied.

2.1. Data Augmentation and SL

Images are the most commonly used machine learning data type for manipulation and artificial sample synthesis. This is because the content of an image can already be sufficiently altered by simple modifications such as adding minor perturbations (e.g. noise, blur or contrast) or minor transformations (e.g. mirroring, rotation, shearing or zooming) (Asperti and Mastronardo, 2017). A recent, more advanced possibility to change data content is the modification of image style with the help of Generative Adversarial Networks (Perez and Wang, 2017). Common to all those manipulations however is that they cannot be applied to time-series data without altering the semantic content of the underlying data: already simple changes like mirroring can render a signed expressions meaningless from a linguistic point of view.

To date, few studies discuss the effect of data warping on machine learning corpora of time-serial data such as handwriting strokes (Wong et al., 2016) or data acquired by an accelerometer (Munoz-Organero and Ruiz-Blazquez, 2017). However, these data streams are of relatively small dimensionality as compared to full body signing movements, and potential data augmentation strategies have to be investigated separately to validate their benefit with a respective SL corpus.

3. Experimentation

In this work, we evaluate the benefit of multiple motion sequence augmentation methods on the base of the performance of a Recurrent Neural Network (RNN) for the recognition of signed sentence content. This RNN constitutes a simple implementation of the sequence to sequence model (Seq2Seq) (Sutskever et al., 2014) for English-French translation with a set of Long-Short Term Memory (LSTM) cells and is similar to a simple encoder-decoder pipeline (Cho et al., 2014). One of the main advantages of this network architecture is that no prior data segmentation is required and the signing sequences can be used as is as network input. The network was trained using a specialized corpus of signed sentence expressions in JSL captured with an optical motion capture system. Here, the idea was to utilize a corpus of data streams that are highly detailed and accurate, so that they could not only be used for the recognition of signed content but also for the generation of virtual signing avatars in the future.

3.1. Corpus and Network Details

379 sentence structures were signed in 2 to 3 different speeds by one fluent signer (Child of Deaf Adults) and simultaneously recorded utilizing a dense Vicon optical motion capture system of 42 cameras (Figure 1). Position and rotations of all relevant body and finger joints as well as facial motion capture were acquired to build a dense set of signing movements able to represent all relevant aspects of a signed expression.

In total, 740 sentences with a vocabulary of 195 words and their corresponding gloss annotations were recorded. Within this corpus, 69 groups of 4 to 6 sentences with similar vocabulary were composed to ensure the repetitive occurrence of the chosen word content. Every group of sentences furthermore contained basic grammatical structures of JSL including non-manual signs and context information (Brock and Nakadai, 2018). In concrete, these were directional and syntactic information such as affirmation, negation and interrogation, adjective inflection (annotated as (CP2) for the comparative and (CP3) for the superlative), logical content separation, as well as compound verbs built from space and size classifiers (annotated e.g. as CL(P) for a location indication or CL(2ppl) for a two-person indication).

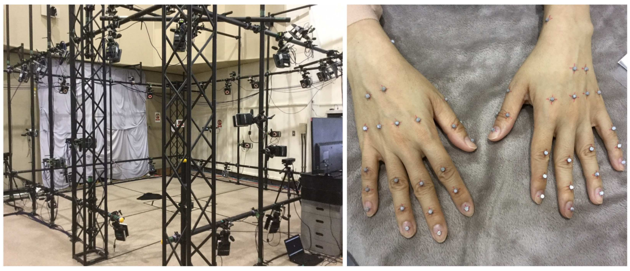


Figure 1: A set of 48 optical cameras was used to record the signed motion sentences. Marker were densely placed on body, finger and face of the sign speaker to obtain a dense collection of sign motion data.

One sentence pattern of each group was randomly chosen for evaluation (154 sentences) and the rest for training (586 sentences). For every data augmentation strategy tested, the number of available training data was then increased by synthesizing new data streams from the original motion capture training data. Next, a new network was learned and the network performance for recognition of the test sentences determined. According to the specification of the Seq2Seq model, the model was trained taking as input the JSL motion sentences and as output one-hot encoded vectors of the respective sentence expression in gloss annotation. To recognize words within the sentence, an additional layer then performed a softmax function on the separate encoding parts (representing classes) of each output sequence and chose the word with the higher probability each time. Lastly, network outputs were post-processed by removing repeated output words.

Throughout the experiment, network parameter such as the number of hidden layers or the size of the LSTM cells were left unchanged to enable the best comparability within the various training results. These parameters were: 1 hidden layer with 256 LSTM cells and six different buckets of length 400, 500, 600, 700, 800 and 900 frames to account for varying sentence length. The recognition network was trained for 2000 epochs for every comparison and the loss function employed was the cross entropy. For the optimization of the LSTM cell parameters, we used the Adagrad optimiser.

It should be noted that motion capture data are seldom used for sign recognition in practice due to the high cost, overhead and effort of data acquisition. Instead, motion captured data sets are more relevant for sign synthesis and avatar animation tasks. For the comparison in this work, we still chose to implement and evaluate a recognition network instead of a generation network, mostly due to the reason that the output of recognition networks are easier to evaluate in terms of accuracy and training loss. Since the Seq2Seq model can be used to generate and at the same time understand sequences, the identical model could also be used for synthesis of a motion sequence from gloss annotation input in the future.

3.2. Applied Augmentation

Three data augmentation strategies were chosen to increase the number of available training data in this work: noise, reversing sequence streams based on anthropometric specifications and sequence warping according to length of different sequences of similar content. Every approach enabled us to double the respective underlying corpus size, leading to differently sized training corpora for the learning of five independent sign recognition networks. These were a simple baseline network trained on only the motion captured data, one network using additional noised corpus data, one network using additional anthropometric reversed sequence streams, one network including warped variations of the original and reversed streams and one network combining all of the augmentation strategies.

3.2.1. Noise

The simplest way of creating new samples must be adding noise to the existing ones. We assessed the effect of two noise types here: basic Gaussian noise and Perlin noise known to better suit the characteristics of human motion.

Gaussian Noise We considered a random variable following a Gaussian distribution $X \sim \mathcal{N}(0, 0.02)$. At each time step, new coordinates were obtained by adding this random variable to the former data:

$$\begin{cases} x' = x + \mathcal{N}(0, 0.02) \\ y' = y + \mathcal{N}(0, 0.02) \\ z' = z + \mathcal{N}(0, 0.02) \\ \alpha' = \alpha + \mathcal{N}(0, 0.02) \\ \beta' = \beta + \mathcal{N}(0, 0.02) \\ \gamma' = \gamma + \mathcal{N}(0, 0.02) \end{cases}$$

with x (x') representing the former (new) position along the lateral axis, y (y') the former (new) roll position along the dorsal axis, z (z') the former (new) position along the vertical axis, α (α') the former (new) pitch angle, β (β') the former (new) roll angle and γ (γ') the former (new) yaw angle. We added this noise to only five significant joints, namely the head, both elbows and both wrists (Guo et al., 2016).

Perlin Noise Different than for the Gaussian noise, noise characteristics were chosen in dependence on the considered joint: the more distal the joint, the lower the noise amplitude and the higher the noise frequency. We used amplitude and frequency values demonstrated to be meaningful for use with SL motion capture data (Mcdonald et al., 2016) to add Perlin noise to hips, waist, upper spine, neck, shoulders, elbows and wrists.

3.2.2. Anthropometric Reversing

In JSL, some words, like the name "Yamamoto" or "book", need both hands to be expressed whereas other ones, like the name "Sato" or "mother", are one-handed. For the latter, both the right and the left hand can be used. JSL signers usually have a "dominant hand": right-handed signers use more often their right hand, and vice-versa. However, it is possible to switch hands if it is more convenient in a particular situation (e.g. when driving or holding something) (Nakamura, 2006), or to emphasize spatial content information within a sentence expression. Therefore, it is critical that our corpus contains both right-handed and left-handed examples of each words. Woefully, the original dataset was strongly imbalanced because data was recorded from only one right-handed signer. That is why additional data representing the same motion as executed by the opposite body half should improve robustness of a training corpus.

To synthesize such movements, we mirrored the data streams of all upper body joints along the vertical axis by swapping right and left values. Moreover, we changed the sign of pitch translations as well as roll and yaw angles. To give a more natural feeling, roll and yaw angles of the head also took their opposite values. Using a virtual avatar displaying the motion capture data (Figure 2), we verified the naturalness and accuracy of the anthropometric reversed sentence expression.

Figure 2: A screen-shot of the original motion (Right) and of the reversed motion (Left), executed by our avatar.

3.2.3. Dynamic Time Warping

For each corpus sentence, at least two repetitions of different signing pace were available, providing the possibility to generate new time variations with Dynamic Time Warping (DTW). First and foremost, it is interesting to note that all recorded data contained static phases at the beginning and at the end of no content information. Therefore, all sentences were first truncated by a simple threshold metric that kept insignificant sequence parts small. The shortest sequence among all captures of identical sentence content was then aligned to the length of the longest sequence, and vice versa.

Let S_1 and S_2 be two sequences. In order to obtain a sequence of the same length as S_2, S_1 is stretched ($S_1 < S_2$) or squeezed ($S_1 > S_2$) by uniform scaling. Scaling is not necessary for sequence alignment, but considered beneficial to increase invariance to large variances in global scale (Fu et al., 2008). The formula to scale a time series $Q = (Q_1, Q_2, \ldots, Q_m)$ of length m to produce a new time series $QP = (QP_1, QP_2, \ldots, QP_p)$ of length p is defined by:

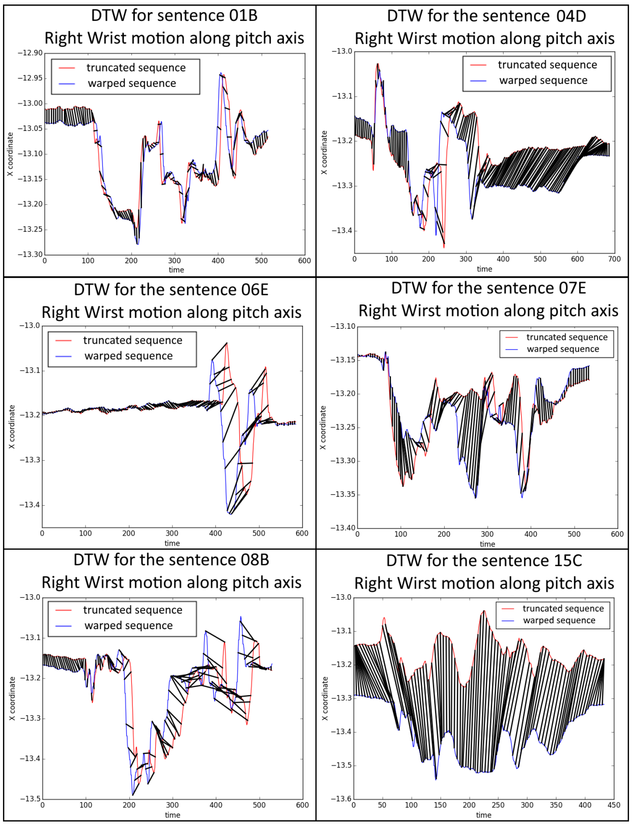


Figure 3: Six sample alignments obtained by applying DTW to uniformly scaled sequences.

$$QP_j = Q_{j\cdot\frac{m}{p}} \quad where \;\; 1 \leq j \leq p$$

Next, DTW was used to adjust the sequence to small local misalignments. We made use of the implementation of fastDTW (Salvador and Chan, 2007) (Figure 3). For two sequences $S_1 = (x_1, \ldots, x_N)$ and $S_2 = (y_1, \ldots, y_N)$, we obtained the optimal warping path P of length L: $P = ((a_1, b_1), \ldots, (a_L, b_L))$. The best alignment (i.e. with minimal total cost) assigned x_{a_ℓ} to y_{b_ℓ} with $\ell \in [\![1, L]\!]$.
The warped sequences showed some defects and appeared less natural than the original data when transferred onto the virtual avatar. On the one hand, we could observe stutters in the motion data generated by expansion of the shorter sequences. On the other hand, transformation of the longer sequences gave rise to hasty motion. As a consequence, warped sentence expressions might be difficult to understand when displayed using the virtual avatar. In order to make the motion more natural, we added a smoothing function to the warped sequence outputs. Clearly, this solution has drawbacks: gestures lost in accuracy and sharpness. Applying a moving average over 5 or 11 data points seemed to be a good middle-of-the-road solution.
We applied this method to both the original and anthropometric reversed data. Thus, we create 2960 additional files for network training.

4. Results

In this section, we present the results obtained with the five different corpora mentioned hereinbefore. For comparison, the evaluation set was always composed of the same

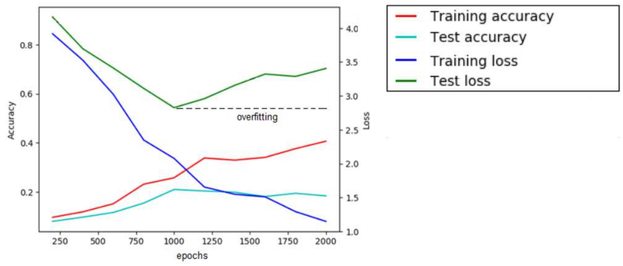


Figure 4: Loss and accuracy evolutions during network training whithout data augmentation, here shown for Bucket 3 of length 600.

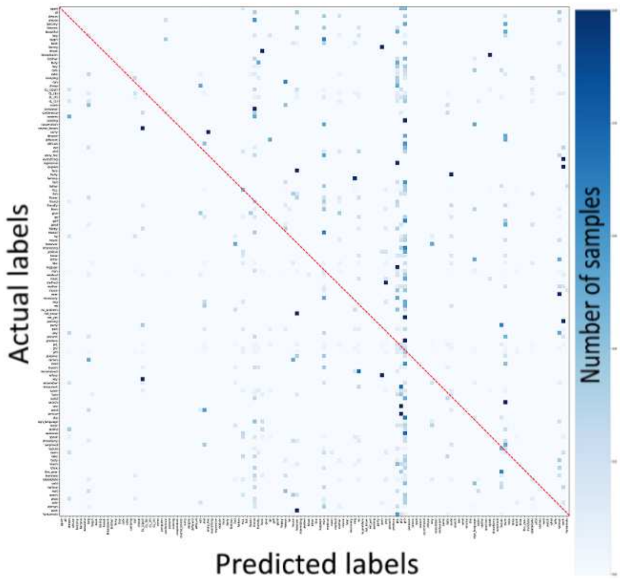


Figure 5: Normalized confusion matrix plot for the sign recognition model trained on non-augmented data and evaluated on the test set.

154 samples describing the 69 different sentence patterns. To speed up the compilation, we did not take into account the facial motion capture data which is commonly also not available for recognition scenarios.

4.1. Original corpus

Network learning on the original corpus only gave insufficient results with test accuracy never exceeding 20%. As Figure 4 illustrates, test loss stopped decreasing after few hundred epochs and even started to increase shortly after. Such training characteristic is commonly known as overfitting, and the resulting model must fail to fit (and hence correctly classify) unknown data. The normalized confusion matrix for the learned recognition model (Figure 5) proves this assumption, as many words were confused with each other. Finally, this model was not efficient with the given recognition task and returned incorrect sentences without meaning (Figure 6).
Both the confusion matrix and the decoded sentence output show that the network model chose certain words more frequently than others. In particular, the words “pt1” (pointing towards oneself), “pt2” (pointing towards a second, opposing person) and “pt3” (pointing towards a direction or object in space) were misclassified and repeated in arbi-

Sentence ID	01E_01
True annotations	pt2 mother pt3 cafe CL(P) pt3 tasty(CP3) banana strawberry cake eat end pt3?
Result	pt1 pt1 pt3 pt3 however mother mother no no pt3 pt3 pt3 pt3 pt3 pt3 pt3 pt3 pt3
Sentence ID	26C_02
True annotations	Sato woman what pt3 translate man skillful despite recommend no pt3
Result	pt2 friend friend always can pt2 swim swim no pt3 pt3 pt3 pt3 pt3 pt3 pt3 pt3 pt3
Sentence ID	69A_03
True annotations	this year Hawaii go think pt1
Result	Suzuki man friend woman CL(2ppl) Hawaii swim past pt1 pt3 pay pt3 pt3 pt3 pt3 pt3 pt3 beg

Figure 6: Three examples of the decoded sentences obtained by evaluating the model trained on non-augmented data.

trary order. All three words correspond to pointing gestures that indicate spatial and contextual references and are frequently used in JSL sentence structures. Indeed, they can occur multiple times within one single, grammatically correct JSL sentence and represent the three most represented words in the corpus. As expected, a simple machine learning corpus of JSL expressions should consequently not be considered sufficient to learn a reliable and meaningful recognition network. On the other hand, lexical items that are of very discriminant structure – such as for example the sign for "Hawaii" – could be retrieved correctly from the test data, indicating that the network model is generally capable to learn the specific features of different lexical items.

4.2. Noise

Gaussian noise The inclusion of training data augmented with Gaussian noise seemed to reduce the overfitting problem as the training loss decreased constantly (Figure 7). However, the test loss still remained quite far from training loss. As compared to Figure 4, test accuracy was brought closer to the training accuracy and now varied between 15% and 35% over the different buckets. Further analysis using the confusion matrix and decoded sentence recognition did not show any improvements and the output could still not be considered relevant or reliable. Semantically (and also morphologically) different lexical items such as for instance given in the sentences 26C_02 and 69A_03 (as listed in Figure 6) could not be distinguished. Moreover, again all pointing gestures were abnormally often detected and predicted labels tended to be selected from a restricted set of vocabulary. We assumed that Gaussian noise-augmented data may either not be different enough from the original data – or too different from the original data, respectively – to significantly improve the learning process. Another possible explanation of those results could be that the five considered joints were not significant enough: in a signed expression, finger joints, facial movements and eye-gaze often carry further meaningful information. We therefore expanded the previous augmentation to all joints, but did not obtain considerably different results. As a conclusion, we did not keep Gaussian noise files in the following.

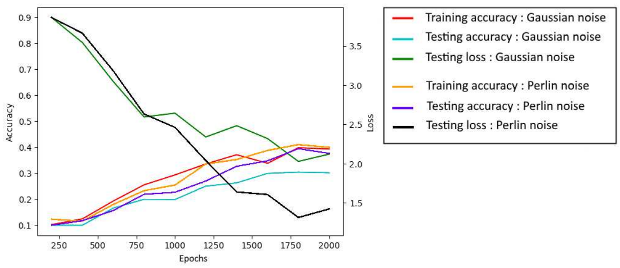


Figure 7: Loss and accuracy evolution during network training when using the original MoCap data and additional noise-augmented data, here shown for Bucket 3 of length 600. Gaussian noises come to the red, green and cyan curves. Perlin noises produce the black, orange and purple curves.

Perlin noise Replacing Gaussian noise with Perlin noise decreased the test loss, improved test accuracy and seemed to settle the overfitting problem well: for all the six buckets, the test loss was continuously and smoothly decreasing, and the test accuracy remained close to the training one (Figure 7). We obtained test accuracies between 20% and 55% within the respective buckets.
While studying confusion matrix plots, we noted that signs were still confused to the most frequent lexical items, and especially the referential pointing "pt3". Considering the procedure of the performed data augmentation, these misclassifications could be explained well with the underlying data structure: as all corpus sentences were uniformly augmented, also their internal word count was equally increased. This means that the absolute count of frequent lexical items expanded as compared to the absolute count of infrequent lexical items. In conclusion, it appears reasonable to apply data augmentation only to those parts of the training data that contain none or few of the most frequent words, or to include and augment single words of low frequency, to better balance the general word distribution in the training corpus.

4.3. Anthropomorphic Reversing

Now, let us have a look at the corpus augmented with the anthropomorphic reversing strategy. Similar as for using Perlin noise, the overfitting problem seemed to be solved well, while test accuracies ranged between 20% and 50% within the respective buckets. Nevertheless, no significant positive changes could be registered in either the confusion matrix nor the decoded sentence recognition output and results shall hence not be further dicussed here.

4.4. Dynamic Time Warping

Inclusion of the warped sentences could not further improve the data and gave similar results than the previous (smaller) corpora with respect to network training parameters. Test accuracies varied between 20% and 50% depending on buckets, while decoded sentences were still as irrelevant as before. Supported by the visual data inspection performed with the virtual avatar, we suppose that DTW did not necessarily preserve all meaningful properties of the signing dynamics. Human gestures follow certain laws related to motor control. DTW may introduce data which

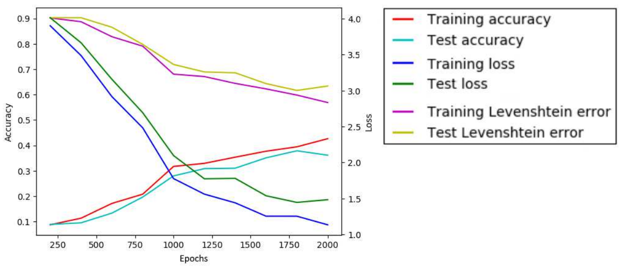


Figure 8: Loss and accuracy evolutions during network training when using the original MoCap data and its anthropomorphically reversed data, here shown for Bucket 3 of length 600.

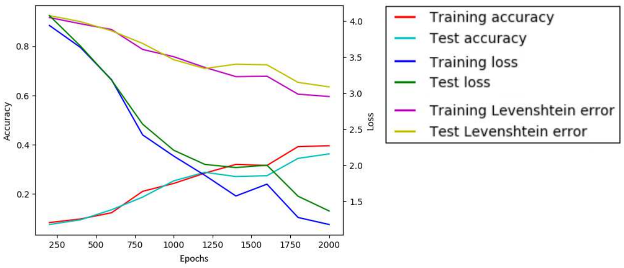


Figure 9: Loss and accuracy evolutions during network training when using the original MoCap data and its warped versions, here shown for Bucket 3 of length 600.

violate these kinematics, leading to unnatural signing data streams that might influence recognition.

4.5. Final set

Finally, we trained our network with a corpus including the original data and synthesized files of all data augmentation strategies. In concrete, these were Perlin noise, anthropometric reversing and DTW applied to both the original and reversed sequences. Resulting test accuracies ranged between 20% and 55% depending on buckets, and good overall network training could be achieved (Figure 10).

Differences between the new confusion matrix (Figure 11) and the original one (Figure 5) are obvious, but oppose our first expectations: instead of higher accurate recognition, signs were even stronger confused to few certain lexical items. Surprisingly, predominantly detected words were not the frequent pointing gestures “pt1”, “pt2” and “pt3”, but words that seem to be randomly chosen from the available corpus vocabulary such as “recommend” and “beautiful”. Moreover, discriminant lexical items like “Hawaii”, whose presence within a sentence pattern could be correctly identified within the test data beforehand, could no longer be retrieved from the test data (Figure 12).

5. Discussion

All of the previous data augmentation strategies improve the general model trainability and extend the amount of corpus data for better use in deep recurrent neural networks, as summed up in Table 1. In particular, the main target of reducing network overfitting could be addressed well by adding a larger number of unknown training data. This is promising as it suggests that further extended data corpora

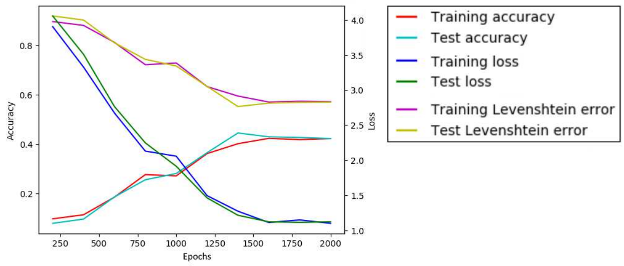


Figure 10: Loss and accuracy evolution during network training when using the final set, here shown for Bucket 3 of length 600.

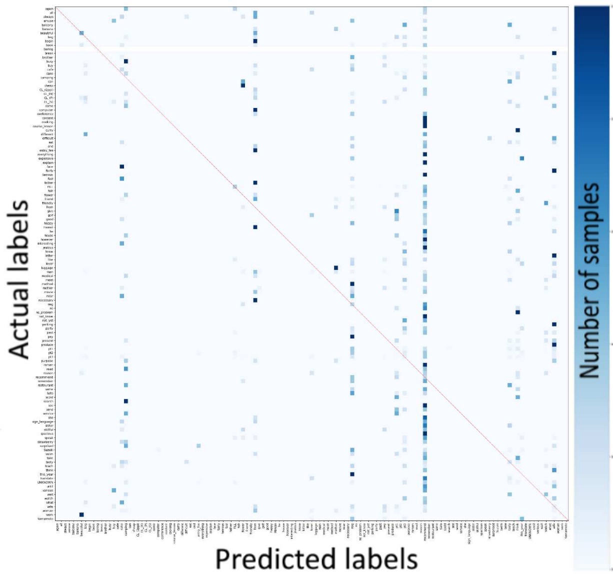


Figure 11: Normalized confusion matrix plot for the sign recognition model trained on the final corpus using all data augmentation methods.

could also improve the robustness and accuracy of future translation systems.

In our specific context, the addition of Gaussian noise may be the less efficient method because general test accuracy does not surpass 35%. Highest test accuracies that could be reached were of approximately 55%. Although this number appears small, results are encouraging for the given context of unsegmented and consecutive SL data streams: to date, continuous sentence expressions have rarely been utilized to learn machine translation networks. Best results were reported for continuous hand-shape recognition using a convolutional neural network trained on a considerably larger set of video data, but could also not surpass accuracies of approximately 60% (Koller et al., 2016).

Despite the improved network trainability, data augmentation could not support reliable sentence recognition. We assume this to be mainly due to the unbalanced characteristic of the training corpus: as discussed, SLs contain specific lexical items that are used repetitively within most sentence patterns. Hence, the network architectures quickly fit to these words and preferably chose the lexical items in their recognition output. Specialized data augmentation strategies such as the Synthetic Minority Over-Sampling

Sentence ID	01E_01
True annotations	pt2 mother pt3 cafe CL(P) pt3 tasty(CP3) banana strawberry cake eat end pt3?
Result	neg pay recommend cake beautiful camping think wife pt3 pt3 recommend recommend pt3 recom. recommend recommend recommend recommend

Sentence ID	26C_02
True annotations	Sato woman what pt3 translate man skillful despite recommend no pt3
Result	neg from recommend lover recom. tasty flower recommend recommend pt3 recommend recom. recommend recommend take take recommend

Sentence ID	69A_03
True annotations	this year Hawaii go think pt1
Result	neg from wife wife recom. watch money pt1 neg recommend recommend recommend pt3 recom. recommend recommend recommend recommend

Figure 12: Three examples of decoded test sentences obtained by evaluating the model trained on the final corpus including all data augmentation methods.

Technique (SMOTE) are shown to improve network performance in imbalanced class problems (Chawla et al., 2002). In the next step, it should therefore be tested whether such data augmentation could also provide benefits for corpora of continuous signed sentence expressions.

To understand the results obtained with the final data set, it is furthermore necessary to investigate why less frequent lexical items were repeatedly misclassified when learning a network using a combination of original data and all data augmentation strategies. Here, it might be possible that the final number of artificially synthesized training samples was too large as compared to the number of available real signing captures, masking out significant features in the training data. In such case, it appears reasonable to introduce a minimum ratio between original and augmented data that should be preserved to ensure successful network training.

6. Conclusion

In this work, we presented and discussed potential data augmentation methods for artificial synthesis of movement sequences that are applicable to the time-serial and complex semantic character of a signed sentence expression. We have seen that the proposed data augmentation strategies are able to increase the number of available training data, while leaving the semantic meaning of the signed expression unchanged. Results show that overfitting, a common problem of small and sparse data sets, can be reduced efficiently. The inclusion of similarly augmented data in any type of SL corpus can therefore be expected to yield better sign language translation networks without the need for additional costly data acquisition or annotation in the future. By removing obstacles of data availability, this could then boost the development of more robust and accurate translation tools.

Data Set	NB Files	Training Acc.	Testing Acc.	Over Fitting	Sign Recog.
Normal	808	0.2	0.1	Yes	Not relevant
		0.3	0.15		
		0.35	0.15		
		0.4	0.2		
		0.5	0.2		
		0.6	0.2		
Noise	1616	0.2	0.2	No	Not relevant
		0.3	0.25		
		0.35	0.3		
		0.4	0.4		
		0.45	0.45		
		0.55	0.55		
Reverse	1616	0.2	0.2	No	Not relevant
		0.3	0.3		
		0.35	0.3		
		0.45	0.4		
		0.45	0.4		
		0.55	0.5		
DTW	2288	0.2	0.15	No	Not relevant
		0.2	0.2		
		0.35	0.25		
		0.4	0.35		
		0.4	0.35		
		0.4	0.4		
All	5384	0.2	0.2	No	Not relevant
		0.35	0.35		
		0.35	0.35		
		0.4	0.4		
		0.5	0.5		
		0.55	0.55		

Table 1: A summary table of improvement statistics. Data set "Normal" is composed of all the original files without data augmentation. "Noise", "Reverse" and "DTW" refer to noise-augmentation using Perlin method, anthropometric reversing augmentation and Dynamic Time Warping augmentation respectively. We gather all augmented data sequences in the data set "All". Columns 2 and 3 (for accuracies) are split into six subsections according to the bucket studied placed in ascending order.

7. References

Asperti, A. and Mastronardo, C. (2017). The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopical images. *arXiv preprint arXiv:1712.03689*.

Brock, H. and Nakadai, K. (2018). Deep jslc: A multimodal corpus collection for data-driven generation of japanese sign language expressions. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, page To appear, may.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Cireşan, D. C., Meier, U., Gambardella, L. M., and Schmidhuber, J. (2010). Deep, big, simple neural nets for handwritten digit recognition. *Neural computation*, 22(12):3207–3220.

Ebling, S. and Huenerfauth, M. (2015). Bridging the gap between sign language machine translation and sign language animation using sequence classification. In *Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*.

Fu, A. W.-c., Keogh, E., Hang Lau, L. Y., and Ratanamahatana, C. A. (2008). Scaling and time warping in time series querying.

Gibet, S., Lefebvre-Albaret, F., Hamon, L., Brun, R., and Turki, A. (2016). Interactive editing in french sign language dedicated to virtual signers: requirements and challenges. *Universal Access in the Information Society*, 15(4):525–539.

Guo, D., Zhou, W., Wang, M., and Li, H. (2016). Sign language recognition based on adaptative hmms with data augmentation.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Koller, O., Forster, J., and Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141:108–125.

Koller, O., Ney, H., and Bowden, R. (2016). Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3793–3802.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. pages 1097–1105.

Mcdonald, J., Wolfe, R., Wilbu, R., and Stec, J. (2016). A new tool to facilitate prosodic analysis of motion capture data and a data- driven technique for the improvement of avatar motion. may.

Munoz-Organero, M. and Ruiz-Blazquez, R. (2017). Time-elastic generative model for acceleration time series in human activity recognition. *Sensors*, 17(2):319.

Nakamura, K. (2006). *Deaf in Japan : Signing and the Politics of Identity*. Cornell University Press.

Ong, S. C. and Ranganath, S. (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):873–891.

Perez, L. and Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.

Pigou, L., Dieleman, S., Kindermans, P.-J., and Schrauwen, B. (2014). Sign language recognition using convolutional neural networks. In *Workshop at the European Conference on Computer Vision*, pages 572–578. Springer.

Salvador, S. and Chan, P. (2007). Fastdtw: Toward accurate dynamic time warping in linear time and space.

Simard, P. Y., Steinkraus, D., Platt, J. C., et al. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *ICDAR*, volume 3, pages 958–962.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.

Umeda, S., Uchida, T., Azuma, M., Miyazaki, T., Kato, N., and Hiruma, N. (2016). Automatic production system of sign language cg animation for meteorological information.

Wong, S. C., Gatt, A., Stamatescu, V., and McDonnell, M. D. (2016). Understanding data augmentation for classification: when to warp? In *Digital Image Computing: Techniques and Applications (DICTA), 2016 International Conference on*, pages 1–6. IEEE.

Communication Across Sensorial Divides – A Proposed Community Sourced Corpus of Everyday Interaction between Deaf Signers and Hearing Nonsigners

Mio Cibulka
Lund University
Helgonabacken 12, SE-223 62 Lund, SWEDEN
cibulka@posteo.de

Abstract

While research on conversation in signed and spoken languages has been flourishing, research on their intersection is scarce. This paper presents an ongoing project that gathers and analyses video data from deaf people's everyday interaction with hearing nonsigners and considers possibilities of involving the communication community that is at its centre and participant empowerment. The scope is to investigate the organisation and structure of communication in which linguistic resources are less accessible and in which social meaning tends to emerge from the interactants' online analysis of the local context (e.g., spatial environment, bodily configurations and movement of the interactants).

Keywords: deaf-hearing interaction, semiotic repertoire, conversation analysis, visible behaviour

1. Introduction

In this paper I present an overview of my research project entitled "A visual ethnography of everyday interaction between deaf signers and hearing nonsigners in Berlin and Tokyo" and contains reflections on ways of involving and empowering those who are at the centre of this research. I hold that the analysis of deaf-hearing interaction offers a great deal of insight into the formation of communicative systems and the emergence and attribution of context-sensitive meaning to bodily visible behaviour. I will furthermore present reflections about this type of communication before going into the matter of how the communication community in question can best be defined and described and how their involvement can contribute to insightful research outcomes.

2. Background

The past decades have seen a rising interest in the lived realities of deaf people. A great deal of research within this sphere focusses on sign language as the main communicative means. This has generated a substantial corpus of studies that are informative of linguistic and communicative features of sign language use. Deaf people, however, predominantly live in societies where spoken languages (both in vocal/oral and written form) are used and where familiarity with sign languages is scarce. For deaf signers, thus, communicating with hearing people who are not familiar with sign language (henceforth hearing nonsigners) is a daily routine, for example in customer interaction.

Interaction between deaf signers and hearing nonsigners (henceforth deaf-hearing interaction) is to date an understudied locus of communication (cf. Kusters, 2017), perhaps because this type of interaction may seem less structured and more improvised compared to forms of communication that primarily employ highly conventionalised, linguistic systems. This research will aid filling this perceptional gap and maintains that gestural/visual communication, regardless of whether it is considered linguistic or not, are orderly occurring forms of communication.

This project is a natural outgrowth of my PhD dissertation (Cibulka 2016) where I studied interactionally relevant phases of manual movement (e.g., hold, retraction and home position) in both signed (in Swedish Sign Language) and spoken conversation (in Japanese, German and Swedish). The upshot was that participants in interaction employ manual movement and nonmovement in accordance with the sequential organisation of the talk, such as signalling imminent speakership, waiting for and acknowledging a reply. Participants do this unrelated to language modalities (i.e. regardless of whether the language used is signed or spoken), to language family and to geographic location (e.g., Sweden, Japan), suggesting a possible universal within human interaction.

3. The Project

My post-doctoral research project is being funded by the Swedish Research Council between 2018 and 2020.

A central premise of the project is that in deaf-hearing interaction both deaf signers and hearing nonsigners are equally creating the progression of interaction by employing their respective communicative repertoires. Since deaf people are encountering this interaction more frequently, they tend to be more of an expert: The communicative repertoires employed by deaf people are a skill acquired through regular contact with hearing nonsigners. Such repertoires contain resources that may be linguistic or nonlinguistic and may be part of one out of several modalities. For this reason, I favour the term *semiotic* repertoire (Kusters, Spotti, Swanwick, & Tapio, 2017) over *linguistic* repertoire. These are developed in deaf interactants through accumulation of interpersonal experiences in their life trajectories (Blommaert & Backus, 2013) and reflect a set of practices that typically and regularly prove successful in interaction with hearing nonsigners.

3.1 Scope

My research interest lies within the emergence of (interactional/social) meaning and the diversity of multilingual and multimodal practices. I intend to look at deaf-hearing interaction with the following questions in mind:

(1) Etic perspective

How much diversity is there in terms of semiotic repertoires with regard to the range of communicative situations? What kind of resources (linguistic and nonlinguistic) are routinely being mobilised for meaning-making and what kind of resources are conferred meaning

to on the spot, especially with regard to the interactants' diverse cultural and linguistic background (in Berlin and Tokyo respectively).
(2) Emic perspective
How do deaf and hearing interactants themselves characterise this type of interaction? To what extent does the interactants' previous exposure to semiotic diversity (e.g., exposure to signed/spoken languages) play a role? How is such interaction socially organised (e.g. in terms of turn-taking) and to what extent do its features overlap with those of other forms of interaction (e.g., purely spoken, purely signed)?
As a whole, this project contributes to recognising and presenting human diversity by highlighting the lived realities of participants in deaf-hearing interaction. It also encourages to think about and to revise traditional concepts of communication, language, sign and gesture, and it contributes to what has been called "comparative semiotics of kinesic expression" (Kendon, 2008). That is, a method of semiotic analysis that does not make an *a priori* distinction whether a given segment of behaviour is sign/gesture, verbal/nonverbal or part of signed/spoken language.
Furthermore, deaf-hearing interaction is apt to exhibit a lesser extent of linguistic resources, compared to interaction between deaf singers or between hearing people with a common language. As such, it offers a less pre-structured and defined array of semiotic typification. If communication in general is to be understood as an embodied experience, this project aids in shedding light at its embodied nature, when linguistic resources are less available.

3.2 Data Collection

Data collection will consist of (1) a pre-structured survey asking the deaf participants about their communicative habits, degree of exposure to signed and spoken languages and other relevant ethnographic background; and (2) video recordings of the deaf participants' daily routines when interacting with hearing nonsigners. My initial idea was to gather these recordings in nonparticipatory observation, i.e. filming without being a (ratified) participant in the interaction myself. I share my reflexions as to how these data could best be gathered in a dedicated section below.

3.3 Analysis

The two types of data will be analysed in two ways: (1) analysis of semiotic repertoires and (2) analysis of action formation and sequential organisation of the interactions with such semiotic repertoires in place. The analysis in (1) is categorical, i.e. I will identify the range of semiotic repertoires in deaf and hearing interactants and associate them with the participants individual background and the larger context (e.g., spatial environment) in which the interaction occurs. The analysis in (2) in is structural, i.e. I will examine the micro-context of semiotic repertoires in a conversation/context analytic (Kendon, 1990, 2004; Schegloff, 2007) fashion. The aim of this analysis is to investigate the sequential organisation of deaf-hearing interaction specifically and how it compares to other types of interaction. See Figure 1 for the study design.

Compared to (purely) signed, spoken or written interaction where interactants may draw from a large pool of conventionalised linguistic resources, interactants in deaf-hearing have a more limited (and more embodied) set of readily deployable communicative resources at their disposal. This leads to a more versatile use of resources (such as objects in the surround or the spatial environment itself; cf. Pennycook and Otsuji [2014]) as social practices. This research is significant, because, as such, offers a novel perspective on the emergence and negotiation of meaning and on the dynamics of meaning-making in human interaction in general.

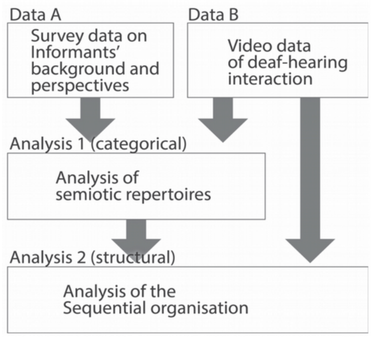


Figure 1: Data and analysis in the project

4. Preliminary Analysis

Some limited amount of data of deaf-hearing interaction were taken at a restaurant in Tokyo (Japan) in 2014 and at a café in Gothenburg (Sweden) in 2016 respectively. In order to give a rough picture of the kind of data that is to be expected, I will illustrate some features of such interaction by presenting two instances of deaf-hearing customer interaction. In both instances we find interactional sequences (namely ordering sequences) that are typical of the respective environment and that are recognised as such, but the resources that participants mobilise are more diverse compared to hearing-hearing interaction.
In the first instance, a deaf customer (C), who is seated at a restaurant table, calls the waiter (W) and checks the availability of a specific dish before ordering. The key frames of this instance are arranged in Figure 2. The duration from the first to the last frame is about 20 seconds. As a general observation, the social relationship between the interactants is established, among others, by virtue of participant mobility: standing-mobile for W vs. steady-seated for C (i.e. a kind of spatial repertoire; described under heading 5). The interaction unfolds within the framework of this social arrangement. C establishes recipiency with W by tapping the wall (creating an audio cue; 2-1) and hand-waving (2-2). When W arrives, C establishes focus on two objects by repeatedly pointing at the notebook screen showing a picture of a dish and at the menu. This way, C establishes a frame for the common activity of identifying an item with the notebook at C's and the menu as W's field of expertise. W engages in this activity by pointing at an item in the menu and gazing to C (2-3). C retracts their hand and creates a hand shape recognisable as OK-sign (2-4). This is a typical slot for placing an order, so the ordering of the item is being entailed by its sequential implicativeness of the micro-context. C then points at the notebook screen again, presumedly showing another picture and shifts gaze to W (2-5). W gives a negative response. The sequence is closed by C smiling, shifting

gaze back to the notebook and changing its orientation away from W (2-6), who shows their understanding of sequence closure by subsequently walking off.

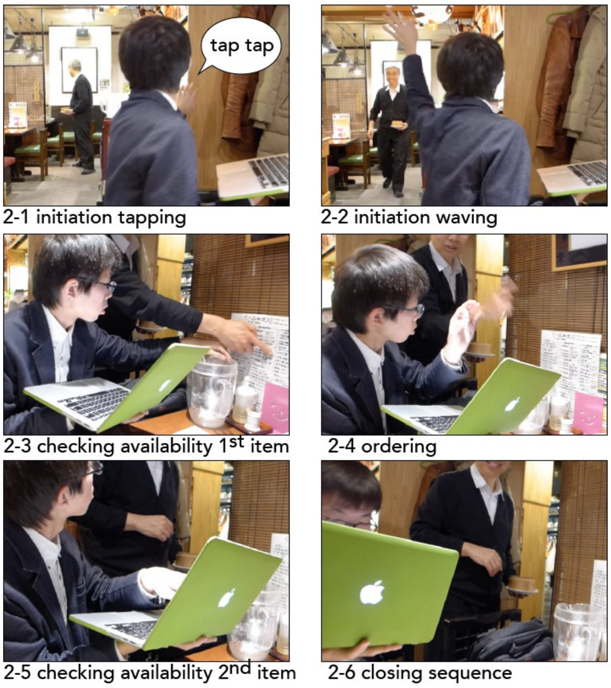


Figure 2: Ordering sequence at a restaurant

A second instance is provided in Figure 3. A deaf participant is standing in line at a café counter where drinks and meals are ordered, served and paid. The social relationship between the interactants is established by the their spatial configuration, where the customer (C) stands in front of the counter and the waiter (W) behind it. The ordering sequence is initiated through C standing in line until W establishes recipiency through eye contact.

In 3-1 C points at a stack of cookies in the back while saying "I want a cookie" in Swedish. W walks towards the stack of cookies, points at it while gazing towards C who nods (3-2). C also orders coffee by saying "coffee" in Swedish (not reflected in the figure). When C presents a banknote, W at first uses vocal resources (3-3) saying that only card payment is accepted. C leans forward and thereby initiates repair (3-4). In response, W points to the card reader (3-5). Understanding is reassured by C by again presenting the banknote while gazing at W who responds with a head-shake and a lateral hand movement (3-6). C displays their understanding by putting the banknote back into the wallet and pulling out a credit card.

Participants thus create focus on objects making relevant specific social actions within the frames provided by the spatial context (e.g., restaurant, queue at the counter) and on basis of known, regularly occurring sequences (e.g., ordering sequences). Taking into account such frames when unpacking social actions is crucial, especially for deaf-hearing interaction, because such frames make relevant and constrict the unfolding of certain interactional trajectories. It is also interesting to examine the kind of resources mobilised by both deaf and hearing participants. For example, C in the second instance mobilises vocal (spoken Swedish) resources on some occasions, whereas in the first instance C relies on auditory (tapping the wall) and visual resources.

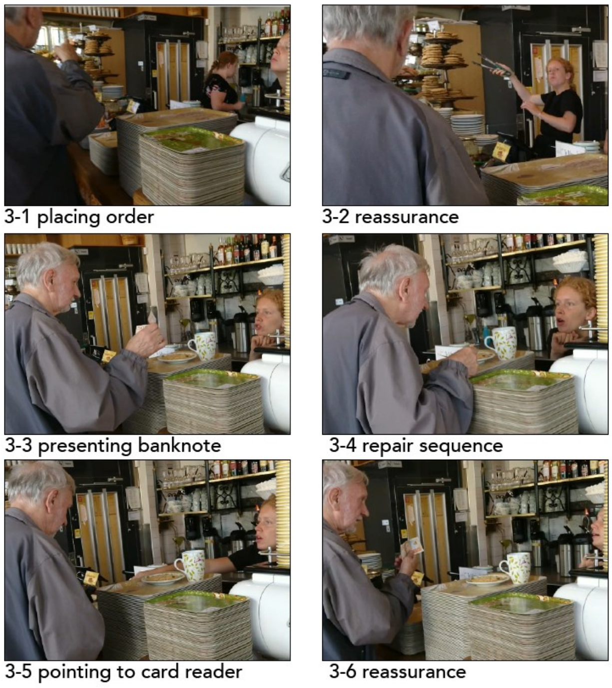


Figure 3: Ordering sequence at a café

5. Language Community

With the overall aims and the general procedure of my project in mind, a rather crucial aspect is who gets to participate. The procedure is rather straightforward for corpora that aim at collecting data in a specific language, as it suffices to recruit participants who are fluent in a given language to the required extent. Deaf-hearing interaction, however, is an intersectional phenomenon; localising the relevant community/communities and defining the focus is a multi-faceted issue.

To begin with, the various repertoires employed in deaf-hearing interaction are not necessarily made up of what formally counts as language: It would be disputable to refer to the communicative system in this type of interaction as "common language" or "lingua franca" between deaf signers and hearing nonsigners. There is thus no communication community for this type of interaction that can readily be called "typical". It is rather described as an intersection between the individual semiotic repertoires of different individuals with asymmetric sensorial access. The very existence of this type of interaction is attributed to the fact that individuals have varying degrees of access to the senses, and that this has an impact on what kind semiotic resources lead to mutual understanding.

Indeed, any type of social interaction can be regarded in terms of an overlap between the participants' semiotic repertoires. In the case of deaf signers and hearing nonsigners this overlap is rather small, when compared to, for instance, that of two fluent signers with a common language. Both interactants do, however, have access to the visual world within reach and understand the contingencies and affordances of the spatial environment in which their bodies are contained, i.e. they have spatial repertoires that may be employed in interaction (see Figure 4).

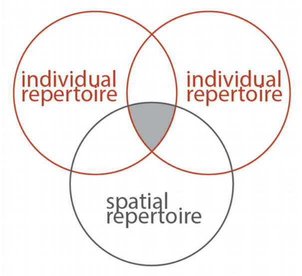


Figure 4: Overlapping individual and spatial repertoires

What is central for my purposes is the way social meaning is made relevant, negotiated and constructed with such individual and spatial repertoires in place, because it requires all parties to act outside the sphere of, foremost linguistic, conventions and dive into a communication of more trial-and-error in order to agree on meaning and thus create conventions and norms that work in the given setting in favour for a shared interactional outcome.

In order to do so, both parties are required to adapt: Deaf interactants may adapt their manual and bodily movement in a way that is understandable to a nonsigner and hearing interactants may adapt their signs/gestures and/or spoken language in a way that is understandable to a deaf person.

Deaf people's lived realities are often embedded in a culturally hearing environment. Thus deaf people may be used to assessing and flexibly adapting to various semiotic resources both in the interlocutor and within the physical surrounds to a higher degree than hearing people who routinely use spoken language (cf. Kusters, 2017).

The ability to assess the kinds of semiotic resource that prove successful in a given interaction is part of deaf people's everyday lives. I regard this an acquired skill and thus an important cornerstone of Deaf epistemology that sit at the intersection of socioculturally hearing epistemologies.

The self-identification of a participant as deaf, hard of hearing, and/or Deaf etc. may be a relevant factor. Broadly speaking, *Deaf* and *Hearing* (uppercase 'D' and 'H') designate cultural, whereas *deaf* and *hearing* (lowercase 'd' and 'h') medical aspects of lived realities related to one's hearing status. The former is a personal choice and way of looking at oneself, the latter is an ascription from a third party.[1]

This study, is neither primarily concerned with cultural aspects of being D/deaf *per se*, nor with cultural aspects of being H/hearing *per se*. The research interest is located within the sphere of sociocultural and communicative diversity that emerges at the intersection of sensorial divides. However, it is important to keep track of the participants' self-identification, since it may offer valuable clues on the choice of semiotic practices. For instance, a deaf participant identifying as culturally Deaf may rely on primarily visible resources, whereas a deaf participant identifying as culturally Hearing may rely on primarily vocal resources.

[1]For a detailed account on D/deaf identities see Padden and Humphries (2006) and Kusters, De Meulder, and O'Brien (2017). See McIlroy and Storbeck (2011) for a discussion on what is called biculturally DeaF identities (i.e. both Deaf and Hearing to some extent).

It has been noted that hearing scholars' epistemological grounds have largely remained undiscussed and that a "productive, (de)constructive exploration of the place of Hearing people within Deaf Studies has yet to occur" (Sutton-Spence & West, 2011). In the light of this discussion it is relevant to clarify my own role as a hearing person who identifies as (culturally) Hearing. I have been in in contact with deaf signers (mainly in academic contexts) and I am somewhat able to to hold a conversation in Swedish Sign Language and to a lesser extent in Japanese Sign Language. This research project is thus an opportunity to put my own background as a hearing person into the equation by scrutinising and presenting the various ways of communication between myself and deaf collaborators.

6. Community Sourcing

A great deal of interaction at shops other service-related businesses happens using spoken languages. This circumstance puts deaf people into the position of being the driving force of deaf-hearing communication in customer interaction. Research into deaf-hearing interaction thus relies on accounts from individuals who feel ostracised from certain life domains because of their hearing status and on video data from this interaction.

The initial idea was that I – as a researcher – conduct fieldwork through following deaf participants through their daily routines in nonparticipatory observation when they interact with hearing individuals. This would give me control over what to record, where to put the focus etc, but on the other hand I would obtain video data that is shaped by my own expectations towards the interaction. An alternative is to leave filming to the participants themselves, as collaborators, and let them decide what, when, how long to film and what phenomena to focus on, i.e. as *participatory* video study.

Researchers observe, analyse and/or represent the lives of others and as such run risk of (re)producing a power relationship with the participants that has been described as "possibly exploitative" (Cunliffe and Karunanayake, 2013). The categories of "researcher" and "researched" are socially constructed roles that may be enacted and reproduced in a variety of ways (Whiting, Symon, Roby, and Chamakiotis, 2016). A part of the research tasks that hitherto have been associated with researchers can be transferred onto the participants. This puts them into a role of co-designers of the research subject and process, and it will be more evident what aspects are most central and important to deaf interactants in a society in which the majority are nonsigners.

This way, at least two layers of data can be obtained: (1) recordings of deaf-hearing interaction itself and (2) the frame that the filming participants produce, i.e. what kind of interactions are being filmed and what is important to interactants.

An excellent example for this type of research approach is the work by Kusters (2015, 2017) on encounters between deaf and hearing people in Mumbai. Filming was done in shops and market places by a team that itself consisted of deaf members. Furthermore, the recordings served as a basis for a full-fledged documentary film, entitled "Ishaare – Gestures and Signs in Mumbai" (Kusters 2015). It contains subtitled interactions, interviews with the deaf protagonists and the shopkeepers. The data from such a large-scale project can be recycled and analysed from a multitude of perspectives: interactional analysis of deaf-hearing encounters; analysis of the way the filming crew orients towards the filming technology and determines the focus of the encounter; reactions from the audience on the film at a screening event, to name a few.

This research frame of participant-based data elicitation contributes to questioning the power relationship between researcher and participant and provides opportunities of participant empowerment.

Digital recording technology has made video making possible for nonprofessionals. The widespread use of smartphones, specifically, enables a great number of individuals to turn a mundane, everyday setting into potential research data, just by tapping the record button. Providing a video camera, a cameraperson and controlled conditions become a less relevant issue for a researcher. What is crucial is to recruit individuals who are willing to register as participants in the project and to put them at the centre. This entails that they are willing to videotape everyday interaction between deaf people and hearing non-signers and/or to have their own interaction with hearing non-signers recorded. Participants are thus assigned both a passive and active role; passive in that their communication is object to enquiry and active in that they are given a free hand over data elicitation and production: the choice over the place, point in time, focus, length of the recording is up to the participants.

For a small-scale pilot project, a website will be set up with information about the background, purposes and procedures about the research project. This will include a registration form for participants, suggestions on how to record and how the data can be transferred. Participants may also categorise, use tags or comment functions to enrich the resulting data; analysis will be a joint process in which participants highlight and/or label phenomena that they ascribe meaning to.

A question that is to be discussed is about procuring the participants' informed consent. This is especially important when collaborators produce video data in public and semi-public spaces such as shops. This is an obstacle because a participant's consent should be obtained before the recording, but on the other hand this task can be documented and studied as a social phenomenon in itself, thus producing another layer of data. That is, both the discussion among collaborators about ways of obtaining informed consent from shopkeepers and the actual outcome of how consent was obtained can be considered valid and central topics of discourse.

7. Bibliographical References

Blommaert, J. and Backus, A. (2013). Superdiverse repertoires and the individual. In W. J. Saint-Georges I. (Ed.), *Multilingualism and Multimodality. The Future of Education Research.* Rotterdam: SensePublishers, pp. 11--32.

Cibulka, P. (2016). *A Universal of Human Interaction? – Manual Movement as Interactional Practice in Spoken and Signed Conversation*. PhD dissertation, University of Gothenburg, Sweden.

Cibulka, P. (2014). When the hands do not go home: A micro-study of the role of gesture phases in sequence suspension and closure. Discourse Studies, 17(1):3–24. https://doi.org/10.1177/1461445614557756

Cibulka, P. (2016). On How to Do Things with Holds: Manual Movement Phases as Part of Interactional Practices in Signed Conversation. Sign Language Studies, 16(4):447–472. https://doi.org/10.1353/sls.2016.0015

Clark, H. H. (2005). Coordinating with each other in a material world. Discourse Studies, 7(4–5), 507–525. https://doi.org/10.1177/1461445605054404

Cunliffe, A. and Karunanayake, G. (2013). Working within hyphen-spaces in ethnographic research: Implications for research identities and practice. *Organizational Research Methods*, 16(3):364–392.

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kita, S., Gijn, I. van, & Hulst, H. van der. (1997). Movement Phases in Signs and Co-Speech Gestures, and Their Transcriptions by Human Coders. In I. Wachsmuth & M. Fröhlich (Eds.), *Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction* (pp. 23–35). London: Springer.

Kusters, A. (2015). Ishaare – Gestures and Signs in Mumbai [Motion Picture]. India: MPI MMG. https://vimeo.com/142245339

Kusters, A. (2017). Gesture-based customer interactions: deaf and hearing Mumbaikars' multimodal and metrolingual practices. *International Journal of Multilingualism*, *14*(3):1–20.

Kusters, A., De Meulder, M., and O'Brien, D. (2017). Innovations in Deaf Studies: Critically Mapping the Field. In A. Kusters, D. O'Brien, & M. De Meulder (Eds.), *Innovations in Deaf Studies: The Role of Deaf Scholars* (pp. 7–52). New York: Oxford University Press.

Kusters, A., Spotti, M., Swanwick, R. and Tapio, E. (2017). Beyond languages, beyond modalities: Transforming the study of semiotic repertoires. *International Journal of Multilingualism*, *14*(3):219–232. https://doi.org/10.1080/14790718.2017.1321651

McIlroy, G., and Storbeck, C. (2011). Development of deaf identity: An ethnographic study. Journal of Deaf Studies and Deaf Education, 16(4):494–511. https://doi.org/10.1093/deafed/enr017

Padden, C., and Humphries, T. (2006). *Inside Deaf Culture.* Cambridge, London: Harvard University Press.

Pennycook, A. and Otsuji, E. (2014). Metrolingual multitasking and spatial repertoires: "Pizza mo two minutes coming." *Journal of Sociolinguistics*, *18*(2):161–184.

Schegloff, E. A. (2007). Sequence Organization in Interaction. Volume 1: A Primer in Conversation Analysis. Cambridge: Cambridge University Press.

Sutton-Spence, R., and West, D. (2011). Negotiating the legacy of hearingness. *Qualitative Inquiry*, 17(5):422–432. https://doi.org/10.1177/1077800411405428

Whiting, R., Symon, G., Roby, H. and Chamakiotis, P. (2016). Who's Behind the Lens? A Reflexive Analysis of Roles in Participatory Video Research. *Organizational Research Methods*, 1–25.

Zeshan, U. (2015). Making meaning: Communication between sign language users without a shared language. Cognitive Linguistics, 26(2):211–260.

https://doi.org/10.1515/cog-2015-0011

SiLOrB and Signotate: A Proposal for Lexicography and Corpus-building via the Transcription, Annotation, and Writing of Signs

Brenda Clark, Greg Clark
brendarc@hawaii.edu, gregclark0130@gmail.com

Abstract
This paper proposes a system of standardized transcription and orthographic representation for sign languages (Sign Language Orthography Builder) with a corresponding text-based corpus-building and annotation tool (Signotate). The transcription system aims to be analogous to IPA in using ASCII characters as a standardized way to represent the phonetic aspects of any sign, and the writing system aims to be transparent and easily readable, using pictographic symbols which combine to create a 'signer' in front of the reader. The proposed software can be used to convert transcriptions to written signs, and to create annotated corpora or lexicons. Its text-based human- and machine-readable format gives a user the ability to search large quantities of data for a variety of features and contributes to sources, such as dictionaries and transcription corpora.

Keywords: transcription, orthography, annotation tools

1. Introduction

This paper proposes a system of standardized transcription and orthographic representation (Sign Language Orthography Builder; SiLOrB), which corresponds to a text-based and searchable dictionary- or corpus-building and annotation tool (Signotate). Though a few notation and transcription systems have been created for sign languages in the past (c.f. Stokoe, 1960; Hanke, 2004, and Sutton, 2010), none have yet become the standard, perhaps due to difficulty of use and structure that does not resemble the structure of signs. Consequently, sign language data is often presented with sign-by-sign or morpheme-by-morpheme glosses as a substitute for phonetic or phonemic transcriptions, even though this would be considered to be unacceptable for spoken language research. It also means that sign language corpora generally cannot be searched on the basis of phonetic aspects and that notes about articulation are not standardized. It is roadblock to lasting documentation and collaborative description.

While increasing use of any type of transcription and line drawings in sign language descriptions, dictionaries, and analyses is certainly a step in the right direction, line drawings and even videos do not indicate the same level of abstraction as a writing system. No matter how close it is to citation form, a line drawing or video clip is one example of a single signer with a specific dialect performing a sign. A writing system like SiLOrB has the potential to consistently represent the phonemic components of a sign with no distractors, and Signotate software allows for easy creation of such representations. Not only is this valuable for linguistic analysis, but it can be used to build easily-searchable and visibly transparent corpora and to create printed literature. While video or live signing is certainly the most reliable mode for communication (just as audio or live speaking is ideal for spoken languages), written language is a way to spread information in communities with limited access to video resources and to allow signers who are hesitant about being recorded to contribute to the conversation.

The sections below discuss the transcription and orthographic conventions used in SiLOrB (2), which aim to improve on existing systems, and the software that can be used to annotate texts with this type of notation (3). Both currently exist as early versions which aim to become a standardized and inclusive system for building sign language corpora.

2. Sign Language Orthography Builder (SiLOrB)

Past attempts at transcription or writing systems for sign languages have had shortcomings such as lack of completeness (e.g. limited non-manuals), use of non-standard characters (as in HamNoSys and SignWriting), and linear organization (as in HamNoSys and Stokoe notation). The SiLOrB transcription system is designed to be analogous to the IPA used for spoken languages: a universal set of ASCII characters and multi-character 'codes' which correspond to individual phonetic components of a language (Clark, 2018). It is transparent, customizable, and creates sign language texts which can be searched based on phonetic or phonemic aspects. The conventions it establishes can be easily incorporated as a system of organization and presentation for lexicons and other texts. It is both machine-readable and human-readable, and has been created with both signers and linguists in mind.

The SiLOrB system also corresponds to pictographic symbols which can be used to create orthographic representations of signs. Each code has a direct and predictable impact on the appearance of a sign's written form, though most symbols are combinatory. For example, a single complex symbol is used to depict the orientation and shape of a hand (see Figures 1-3). This type of symbology allows for a less linear representation which more accurately reflects a sign's articulation and phonemic structure. The sections below discuss basic aspects of the transcription (2.1) and writing (2.2) systems. Those who are interested in the full current version can visit https://bleegiimuusclark.com/home/silorb-sign-language-writing/.

2.1 Transcription

The system described here is based on what is known about phonemic distinctions in sign languages (see Jepsen et. al., 2015; Crasborn et. al., 2000), and aims to improve on existing forms of transcription. An early version of SiLOrB was used to write Sivia Sign Language (Clark, 2017), and the current version (2.0) is expanded based on additional phonological distinctions used in ASL, Hawai‘i Sign Language, and a few others. SiLOrB breaks a sign's articulation into the familiar categories of 1) shape and orientation of the hands, 2) location, 3) movement, and 4) non-manuals. Each category is specified with a capital letter code followed by the applicable phonetic information in a set order. A user can choose to describe a sign at the level of detail necessary for their objective, in order to fit the language’s phonology or morphology, or even to record contrasting phonemic and surface forms of a sign. Because categories and descriptors are additive rather than mandatory, morphemes consisting of fewer components, such as a modification to handshape, a type of movement, or a facial expression, can be depicted individually as well.

After the specification of the dominant or non-dominant hand (D or ND), palm and finger orientation are given, followed by groups of fingers and their positions. ‘DND^Vc*%A^+’, for example, means that both hands (DND) are in a palms upward (^), fingers forward (V) position with all the fingers rounded (c) and making contact (*). Then both hands change (%) to palms bodyward (A), fingers upward (^) orientation with all the fingers extended (+). Locations (L) consist of a regional code and optional further specification of placement and contact, as in ‘Lzv>< o’ describing a location in zero space (z) below the waist (v) and near the vertical center (><), which is close to the torso but does not make contact (o). Movements (M) often consist of a direction and a path, as in ‘M^sm’ for upward movement (^) with a short trajectory (sm). Non-manuals (NM) give a part of the body followed by a position or movement code. ‘NMM*’ describes the mouth (M) in a pursed position. Thus, the sign for ‘fire’ in Sivia Sign Language (LSSiv) is transcribed as ‘DND^Vc*%A^+; Lzv o; M^sm; NMM*’ (see Figure 1).

For simpler signs which do not utilize every aspect, unnecessary categories are simply deleted from the transcription. The LSSiv sign for Peru, for example, uses only the dominant hand (D) in a consistent orientation and shape: forward palm (V) and upward fingers (^) with the index (1) and middle (2) spread (w). Its location (L) is simply the forehead (fh) with contact (x), and there are no movement or non-manual components. Thus, the transcription for ‘Peru’ is ‘DV^12w; Lfhx’.

SiLOrB transcription has also changed from its original version to use more iconic coding conventions which limit language barriers for users who are not fluent in English. Though top-level category codes and some specifiers are based on English terms, arrow-like characters (^, v, <>, ><, >>, <<, A, V) are used for orientation, movement, and location specifiers. Similarly, emoticons are the inspiration for many mouth shapes, such as ‘)’ for a smile and ‘P’ for an exposed tongue. Other codes are chosen to resemble a corresponding orthographic symbol, such as ‘%’ for a change in position (as in Figure 1) or ‘*’ for contact made with the fingertips.

Distinctions such as ‘in’ (toward the center) versus ‘toward the dominant side’ are clarified using digraphs (>< versus <<). Instead of requiring the absolute direction of each hand, ‘inward’ and ‘outward’ options allow a user to reference a vertical center line. This creates a distinction between mirrored and purely directional movement, and allows both types to be described with the same value for both hands (e.g. both hands inward or both hands toward the dominant side). This is one of the ways SiLOrB is geared toward phonemic representation and searchability, along with its hierarchical organization.

2.2 Written Representations

As with transcription, the objectives for the writing system are clarity, consistency, and ease of use. Some conventions are inspired by existing systems, such as the use of white for the palm and black for the back of the hand, as in SignWriting (Sutton 2010). The basic structure combines SignWriting’s ‘drawing of a signer’ approach with some linear elements which add to its consistency and readability for longer texts. SiLOrB is described as ‘non-linear’ in contrast to systems like HamNoSys or Stokoe notation, which simply list handshape, orientation, etc. from the left to right. SiLOrB instead uses many combinatory symbols and works largely from the center outward. It does have linear components, however, due to handshape changes which are listed from left to right on both sides, the depiction of non-manuals on the far right, and formatting which standardizes the height of each component to resemble a line of text.

Full orthographic representations of signs (created with Signotate; see Section 3) combine pictographic symbols which are arranged as a ‘signer’ facing the reader. A location symbol is placed in the center with the hands (a combined shape and orientation symbol) on either side and movement to the outside of each hand. Non-manuals are given with the location if applicable (e.g. a central torso image may have markers for hands making contact with that location and for movement of the torso itself), and additional non-manuals occurring on other parts of the body are represented on the far right. Figure 1 shows the orthographic representation of the Sivia Sign Language (LSSiv) sign for ‘fire’ described in the previous section: 1) the center drawing of a torso indicates zero space and circles show that the hands are near a low and central part of the torso; 2) the hand symbols on either side show a change from palm up with fingers in a rounded position to palm bodyward with fingers extended, 3) movement arrows depict a short upward path, and 4) the face on the right edge shows pursed lips. (See B. Clark 2018 for a full description of the current system and examples with corresponding videos.)

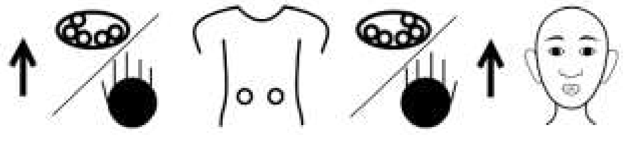


Figure 1: Written representation of the LSSiv sign for ‘fire’ (DND^Vc*%A^+; Lzv>< o; M^sm; NMM*)

As in the transcription system, simpler signs may not use all of the available parameters, and appear with fewer symbols. The sign for 'coca' uses only the dominant hand and has no movement, so its orthographic representation is much shorter, as seen in Figure 2. The sign consists of an extended index finger touching a puffed out cheek. This is also an example of symbology that combines locative and non-manual aspects in a single region (here, the head).

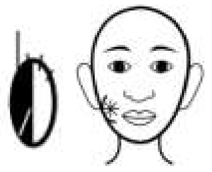


Figure 2: Written representation of the LSSiv sign for 'coca' (D><^1+; Lchk*; NMchk<>d)

Figure 3 shows the sign for 'cacao', which uses the non-dominant hand as its location and only moves on one side. The dominant hand with palm down and bent fingers moves outward repeatedly over the non-dominant hand with palm up and extended fingers. Again, a single unit depicts the features of the non-dominant hand and serves as the location for the dominant hand.

Figure 3: Written representation of the LSSiv sign for 'cacao' (Dv><r ND^><+; Lnd; M<>#)

3. Signotate Software

Software called Signotate is currently being developed to create written signs based on their SiLOrB transcriptions. It is also a tool for creating documents such as a transcript or a lexicon consisting of many annotated signs. Like SiLOrB, Signotate is designed to be intuitive for a variety of users and customizable for a variety of tasks. The following sections discuss specific features of sign entry (3.1) and search functions (3.2). Those who are interested in the project can visit https://github.com/Signotate to find out more and keep up with the latest updates.

3.1 Sign Entry

While SiLOrB transcription code and orthographic symbols can be created by hand as video annotations or entries in a lexicon, the Signotate software application provides an easy way to convert transcriptions into written signs and to create a lexicon or a transcript from multiple entries. It facilitates quick transcription in the field, allowing transcribers to rapidly add collected data to a corpus. Figure 4 shows the application's interface, which guides a user to enter a sign's transcription in the four main categories of hand, location, movement, and non-manuals. The default form is a one-handed sign which occurs on the dominant side of the body, though a user can choose to switch to the non-dominant side. For two-handed symmetrical signs, a 'dual-sided' option automatically copies a transcription to the non-dominant side as well, and for asymmetrical signs, an 'asymmetrical' option allows a user to edit both sides individually.

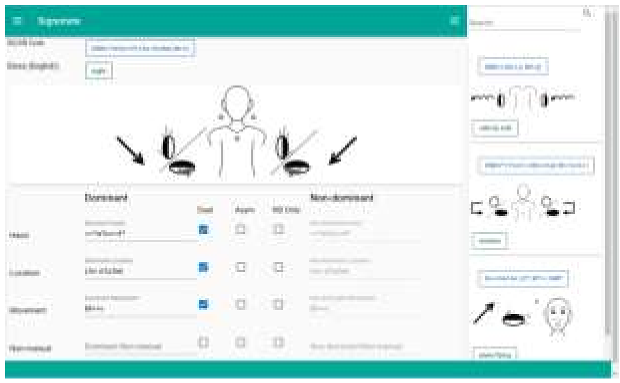


Figure 4: Screenshot of Signotate

Symbols appear in the upper box as codes are entered below to help ensure that the desired configuration is achieved. Additional signs in the same transcript or lexicon appear on the right panel with their transcriptions and glosses, and metadata can be entered below transcription codes. These fields allow a user to follow glossing and annotation conventions such as those outlined in Crasborn, Bank, & Cormier 2015, with the addition of standardized phonetic notation. Future implementations of Signotate may also include plugins for programs such as iLex or ELAN to allow written signs to appear along with time-aligned transcriptions.

3.2 Searchability

The system is indexed by articulatory features that a user inputs, so a Signotate corpus is instantly searchable by phonetic or phonemic components. Aspects like one- or two-handedness and symmetry or asymmetry are also included in searchable components, as well as some implications which are not explicitly expressed in transcription. For example, fingers described as 'bent' are also marked as extended, though SiLOrB coding only requires that 'bent' (r) be specified. Any field in metadata (e.g. glosses, morphemes, participants, location, timestamp, etc.) or code in the form of a sign (e.g. extended fingers, location, type of contact, movement direction, eye gaze, brow position, etc.) is searchable as well. Signotate is also able to perform SQL style searches. For example, one could search for signs that begin at any location below the waist, with the fingers oriented upward. Similarly, one could search for transcriptions that involve a person from Cusco who is between 25 and 30 years of age, and is not a native user of Peruvian Sign Language.

Signotate exists as both a web implementation and standalone desktop implementation. The desktop version, which stores its data locally in an SQLite database, can be used offline, while still enabling search across small to moderately-sized corpora. The web implementation, which is backed by Elasticsearch, is capable of searching across very large corpora. In the future, Signotate could support more complex searches and aggregations, such as phrasal search, or searches for grammatical or syntactic patterns. Signotate lexicons and transcripts can be imported from and exported to a human-readable yaml formatted file, as shown in Figure 5. (See https://bleegiimuusclark.com/signotate-v0-1-yaml/ for a complete example of this format.)

```yaml
transcript:
- times:
    start: 1.51.416
    stop: 1.51.897
  usage: "typical"
  transcriptGlosses:
  - language: "English"
    gloss: "bee"
    morphStruct: "bee"
    signClass: "noun"
    usage: "typical"
  otherGlosses:
  - language: "English"
    label: "meaning1"
    gloss: "sting"
    morphStruct: "sting"
    signClass: "verb/noun"
    usage: "typical"
  sign:
    silorbCode: "D><^t*; Lchk; Mx#"
    properties:
      singleSided: "dominant"
      nonManual: false
    hands:
      silorbCode: "D><^t*"
      dominant:
        silorbCode: "><^t*"
        singleSided: "dominant"
        stages:
        - stageCode: "D><^t*"
          features:
          - palmIn
          - fingersUp
          fingerFeatures:
          - fingers:
            - 0
            - 1
            - 2
            - 3
            - 4
            features:
            - extended
            - tapered
            - contact
      nonDominant:
        singleSided: "dominant"
    location:
      silorbCode: "Lchk"
      dominant:
        silorbCode: "chk"
        singleSided: "dominant"
        stages:
        - stageCode: "chk"
          region: "chk"
          proximity: "near"
      nonDominant:
        singleSided: "dominant"
    movement:
      silorbCode: "Mx#"
      dominant:
        silorbCode: "x#"
        singleSided: "dominant"
        stages:
        - stageCode: "x#"
          other:
          - tap
          - repeated
      nonDominant:
        singleSided: "dominant"
```


Figure 5: Signotate yaml snippet for one sign

4. Conclusion

The culmination of SiLOrB and Signotate is the ability to build corpora of sign languages which include not only glosses and translations for videos, but annotation at several levels, including anything from phonetic features, morphemes, and single signs to extended texts such as narratives or conversations. The resulting corpora would utilize a detailed and universal format for talking about sign languages which is machine readable and easily used by language researchers and computational linguists for a variety of tasks including automated sign language transcription or analysis.

Descriptive, searchable, and standardized annotations combined with Signotate software open the door to new collaborative possibilities for sign linguists, for natural language processing researchers, and for signing communities. Aside from its descriptive and analytical advantages, these solutions will enable a user to create typed, alphabetized, and printed media for sign languages. Not only is this important for data preservation, presentation, and organization, but it can provide options for communities with limited access to video and online resources. The idiomatic nature, extensibility through documentation, and software suite help ensure high quality and long lasting documentation of sign languages for a variety of purposes.

Bibliographical References

Clark, B. (2018). Sign Language Orthography Builder. Retrieved from https://bleegiimuusclark.com/home/silorb-sign-language-writing/

Clark, B. (2017). A grammatical sketch of Sivia Sign Language (Doctoral dissertation). University of Hawai‘i at Mānoa.

Crasborn, O., Bank, R., and Cormier, K. (2015). *Digging into Signs: Towards a gloss annotation standard for sign language corpora.* Nijmegen: Radboud University & London: University College London.

Crasborn, O., van der Hulst, H. and van de Kooij, E. (2000). Phonetic and phonological distinctions in sign languages. In *A paper presented at Intersign: Workshop (Vol. 2)*.

Hanke, T. (2004). HamNoSys-representing sign language data in language resources and language processing contexts. In *LREC, 4*.

Jepsen, J., De Clerck, G., Lutalo-Kiingi, S., and McGregor, W. B. (Eds). (2015). *Sign languages of the world: a comparative handbook.* Walter de Gruyter.

Stokoe, W. C. Jr. (1960). *Sign Language Structure: An outline of the visual communication systems of the American deaf.* In Studies in Linguistics, Occasional Papers, 8. Department of Anthropology and Linguistics, University of Buffalo Buffalo, NY.

Sutton, V. (2010). Sutton's SymbolBank: International SignWriting Alphabet. http://movementwriting.org

Terminology Enrichment through Crowd Sourcing at PYLES Platform

Eleni Efthimiou[1], Stavroula-Evita Fotinea[1], Panos Kakoulidis[2], Theodoros Goulas[1]

[1]Institute for Language and Speech Processing (ILSP)/R.C. "Athena", [2]Innoetics
[1/2]Artemidos 6 & Epidavrou, Maroussi, 15125 Athens, Greece
E-mail: {eleni_e, evita, tgoulas}@ilsp.gr, pk88gr@hotmail.com

Abstract

Here we present how Greek Sign Language (GSL) has provided content to various functionalities of the Information System PYLES, a management system for on-line lessons, designed to support accessible asynchronous e-learning by addressing learning needs of students with various communication capabilities and needs, at the Technological Educational Institute of Athens (TEI-A). In order to meet native signers' student needs, the platform has incorporated descriptions in GSL of various lessons in the curriculum, as well of administrative forms and documents. However, the most useful tool for the student community has been a bilingual on-line lexicon which provides both general purpose content and domain specific terminology glossaries. The major characteristic of this tool is that it allows for content enrichment via crowd sourcing, which, hopefully, proves to provide a satisfactory solution in respect to terminology gaps often noticed one major difficulty in of deaf individuals' education.

Keywords: SL terminology glossaries, SL lexicon resources, GSL-Greek bilingual resources, deaf education services, educational platform accessibility services, terminology crowd sourcing

1. Introduction

In many occasions there has been made reference to the difficulty deaf individuals are confronted with when they need to access a written text (Chamberlain et al., 1999; Leigh & Andrews, 2017; Ghari, 2016). This phenomenon becomes more acute in the context of Higher Education, where students are constantly in contact with written material, very often even written in another language than the one of their hearing environment.

Furthermore, terminology related issues are known to be observed in all cases, where a new term or a set of new terms created within a specific language system, need to be transferred to some hosting language system.

Although this problem has been well observed and transfer procedures are described in detail in a series of ISO recommendations (ISO/R 704, 1968; ISO/R 1087, 1969) and their updates (ISO 704, 1987; ISO 1087, 1990; ISO 1087/1, 2000) since mid 90's, the whole process of terminology transfer and validation is a difficult one and very few terms end up to be widely incorporated in the linguistic reality of the active language domain users.

This paper presents how Greek Sign Language (GSL) is incorporated in the educational material provided to deaf students of the Technological Educational Institute of Athens (TEI-A) via PYLES Information System, a management system for on-line lessons, designed to support asynchronous e-learning.

Besides the use of GSL video for the presentation of administrative forms and documents, the following sections will deal with demonstrating how accessibility to educational content is supported, focusing on an on-line lexicon that encounters a general purpose bilingual dictionary for the language pair GSL-Greek and a number of terminology glossaries from the scientific domains taught at TEI-A at an open environment that allows for contributions from different categories of users in a crowd sourcing appeal, targeting to resources enrichment, following a validation protocol that ensures quality control of the newly added lexical/ terminological items.

2. PYLES Information System Architecture

The Information System PYLES is a management system for on-line lessons, designed to support accessible asynchronous e-learning, addressing learning needs of students with various communication capabilities and needs at the Technological Educational Institute of Athens (TEI-A). It, thus, exploits both up-to-date assistive technology software and content presentation in various forms.

The platform has been used as the basis for the development of an active repository of multimodal educational resources, also incorporating a terminology section that entails domain specific terminology glossaries and a general purpose bilingual dictionary for the language pair GSL-Greek.

The platform provides advanced customization options according to user needs (Fig. 1) but also a collaborative environment for the support of teaching and learning processes (Šumak et al., 2011).

The information system [1] is built on the open code platform *Open eClass*[2], a free e-learning platform that has actually been enriched with tools and functionalities which allow extended accessibility regarding both the environment and the educational content (Efthimiou et al., 2015).

The GSL terminology environment allows for the creation of different glossaries directly by their users, where GSL signers are invited to upload their suggestions for various terms under specific quality control conditions as presented in section 2.2.

[1] http://eclassamea.teiath.gr /.
[2] http://www.openeclass.org/.

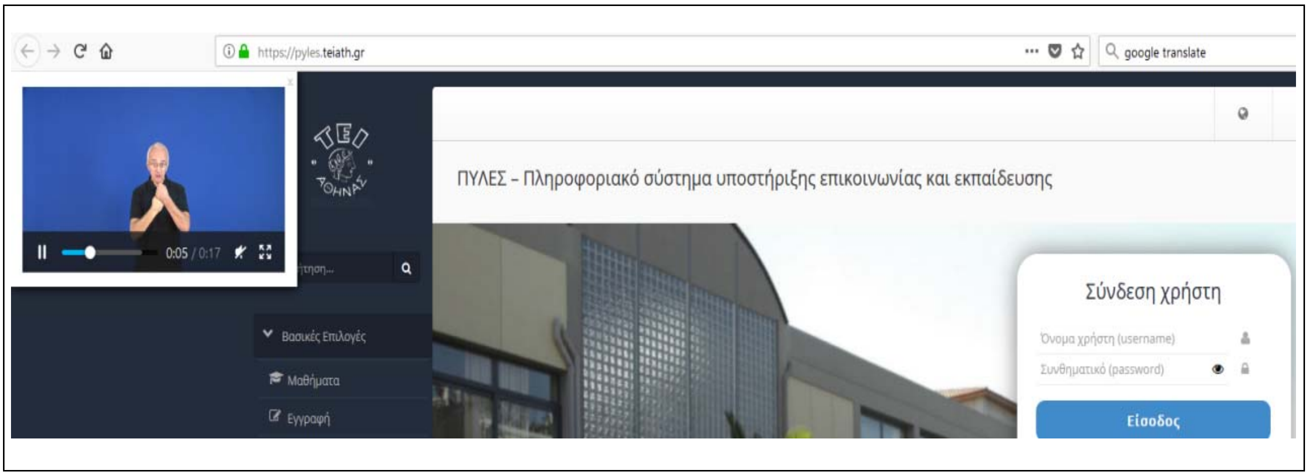


Figure 1: PYLES platform main page with activated navigation information in GSL and user login window.

Two types of terminology resources are available to all users:

a) Terms created and validated according to ISO provisions (ISO 704, 1987; ISO 1087, 1990; ISO 1087/1, 2000),
b) Terms collected via crowd sourcing.

For the search and presentation of lexical items (general purpose lexicon or terminology), the user, after logging in, may select the domain of his/her interest from the list in (Fig. 2) and then typing his/her wished item in the search box.

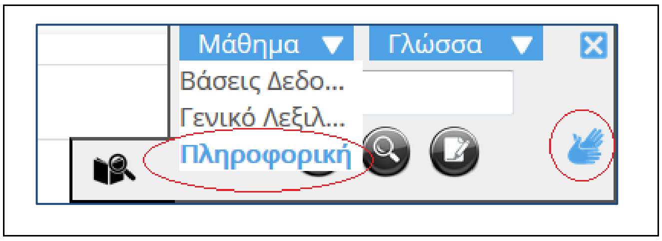


Figure 2: Selection of terminology domain from a list for the search of a term, when GSL presentation mode is activated.

Another option is to search an item from an alphabetically ordered list (Fig. 3).

When in alphabetical search environment, the user is provided also with various statistics such as "latest entrances" and "highly scored entries", along with other crowd sourcing and lemma related information such as the language domain the lemma belongs to, the date of its addition and the language(s) in which it is available.

Accessibility of content at PYLES has been supported by exploiting language based assistive technologies which involve implementation of a synthetic voice facility for the accessing of written content across platform by users with vision problems, along with incorporation of the GSL tools presented here to support GSL signers use the platform.

Activation of all accessibility facilities is subject to customization, according to specific user needs and preferences.

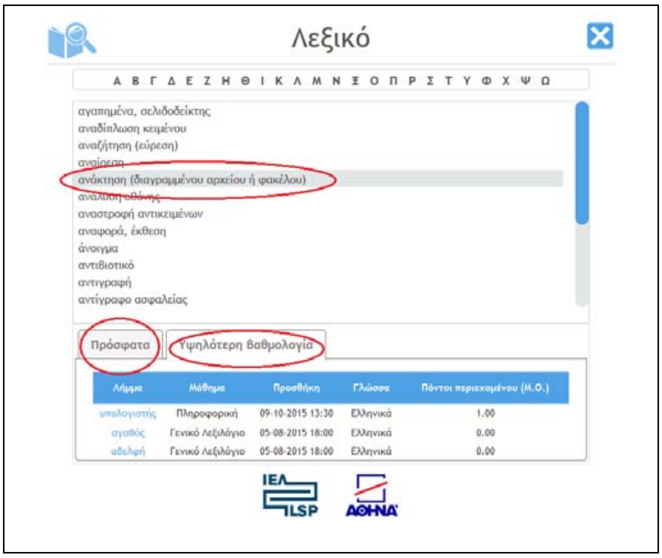


Figure 3: Alphabetical search option.

Regarding overall customization options intended to serve GSL signers' needs (Dimou et al., 2014), the platform incorporates:

- Selected lesson presentations in GSL on the basis of deaf students' preferences regarding the curriculum offers,
- An on line bilingual dictionary of general purpose lemmas for the language pair GSL-Greek,
- Online terminology glossaries which provide terminology items presentation options in GSL, Greek and English,
- Administrative form related information in GSL.

3. Terminology Resources at PYLES Platform

Conceptualization of terminology is a major factor for acquiring new knowledge and also a prerequisite for the production of new knowledge within a scientific domain (Sager, 1994). In this respect, terminology is of major importance in the framework of scientific education and vocational training.
However, when investigating the availability of terminology in various SL national systems, one discovers a critical gap that directly affects deaf students' integration to productive higher education communities.
This situation is due to a twofold problem. On the one hand, introduction of new terminology lists to a linguistic system by following international standards is a difficult and time consuming task that requires dedicated involvement of relatively big groups of experts for a considerably long period. On the other hand, deaf students/scientists usually need to acquire the terminology of their domain of expertise by accessing written texts in English. This task sets an extra burden for everyone been raised in a non English speaking environment, in addition to the generally recognized difficulty a considerable number of deaf individuals face with accessing written information.
At PYLES platform, since the major goal has been to support educational activity, it has been necessary to cope with the terminology lack problem. Thus, two options of terminology list creation have been adopted: (a) by gathering existing term collections created according to ISO recommendations, and (b) by provision for a crowd sourcing activity in respect collecting not yet fully validated (*ad-hoc*) terms, which, however, are actively in use within domain specific user communities.

3.1 Terminology Creation According to ISO Recommendations

The procedure, as foreseen by the various ISO recommendations and their updates relating to introduction of existing terminology into a new language, involves a series of steps associated with different tasks performed by working groups of experts and groups of members from the receiver language community.

More specifically, in *step 1*, a group of experts, composed by individuals with long experience of work within a given domain, examine the related terminology glossary and propose a translation for each new terminology item. This translation must be representative of the conveyed knowledge, be accepted by the technical user community, but also it must follow the rules of the receiver language as regards new lexical item formation mechanisms (Gruber, 1993; Katsoyannou & Efthimiou, 2004). These requirements make it necessary to include theoretical linguists and lexicographers to terminology translation proposing groups.
In *step 2*, the list with the experts' proposal is sent to an extensive working group consisting of representatives of user communities, in order to be checked. This second working group is obliged to return the list with markings for "acceptable", "rejected", "needs correction", "new proposal" for each terminology item undergoing the validation process.

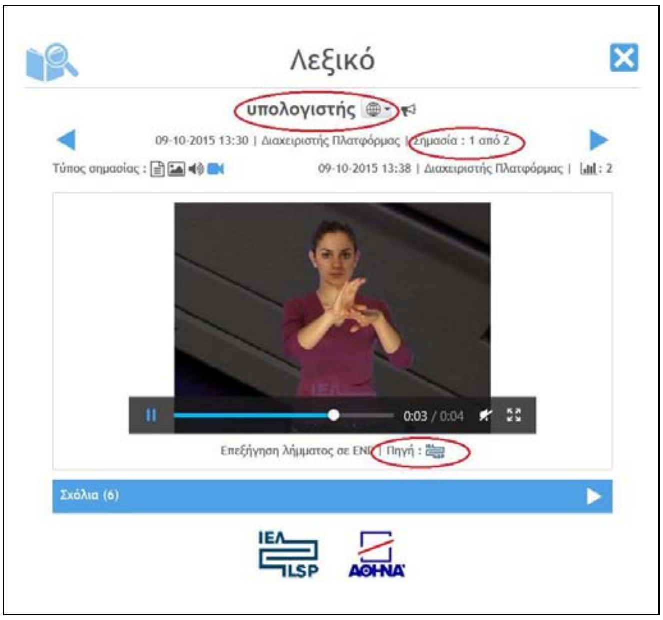


Figure 4: Term presentation window at TEI-A platform.

In *step 3*, the proposing group of experts incorporates the user community's suggestions and delivers the term list again for a second validation process.
Step 4 is the final state of a term validation process, if the user community group accepts the translation as an integral part of their language.
Steps 2 to 4 may be iteratively repeated as many times as needed in order for a term translation to be validated and accepted.
As regards terminology in the TEI-A platform, the glossary of basic computer use terms has been translated to GSL following the above described procedure, where initial suggestions have been gathered from deaf experts teaching *Introduction to Computer Use* classes for several years. Their initial video recorded glossaries have been discussed among members of native GSL signers' community for a long time and in a number of iterative sessions for steps 2 to 4 of the validation procedure before unanimous agreement was succeeded and the term collection was acquired in its final form (DIOLCOS, 2006).
But terminology items, in order to serve knowledge transfer, need also to be associated with a number of other types of information which clarify their meaning within a specific domain of use. Such information incorporates the term equivalent in other languages, a term definition and explanatory visual material like pictures, diagrams, drawings, etc. Of equal significance is reference to the source of information accompanying each term.
For the creation of the terminology presentation window at the TEI-A platform, we have taken all these parameters into account, emphasizing on the platform's educational dimension (Efthimiou & Fotinea, 2007; Efthimiou, 2008;

Fotinea & Efthimiou, 2009; Fotinea et al., 2012). All information display options relating to a specific term that are available to the user, are depicted in Figures 4 to 6. In (Fig. 4) the encircled information accompanying the term display in GSL refers to (i) the Greek equivalent term and the language button, presenting the term equivalent in more languages as i.e. English, both centered at the top of the window, (ii) the senses associated with the displayed form in the next line, and (iii) the source database of the displayed information underneath the video window, which in this case is the term "*computer*" in GSL and derives from the ILSP repository as indicated by the institute's logo.

In (Fig. 5), the definition of the term is the main displayed element. In this case, the source of the displayed information is the Greek Wikipedia, while other displayed information identifies the person who entered the definition and the time of creation of the displayed information.

Figure 5: Term definition presentation. Encircled is information about time of make and creator of entry, as well as its source.

The encircled icons at the top left side of the window indicate the options for the means of information display related to the chosen term, which are available to the user, where blue color indicates the currently chosen option.

In (Fig. 6), the display option is selected, which provides visual support material for the comprehension of the term. Furthermore, the "more languages" button (globe icon) placed next to the Greek term equivalent is also selected, displaying the term equivalent in English. The at least trilingual association of each term (GSL, Greek and English) has been considered necessary, since students at Greek universities very often need to consult bibliography written in English language. In the (Fig. 4) example the central explanatory icon source is again the Greek Wikipedia.

This terminology collection is also used as a pilot for the validation of terminology collections in other scientific domains when translated to GSL, since crowd sourced terms are in most cases characterized as ad hoc terms.

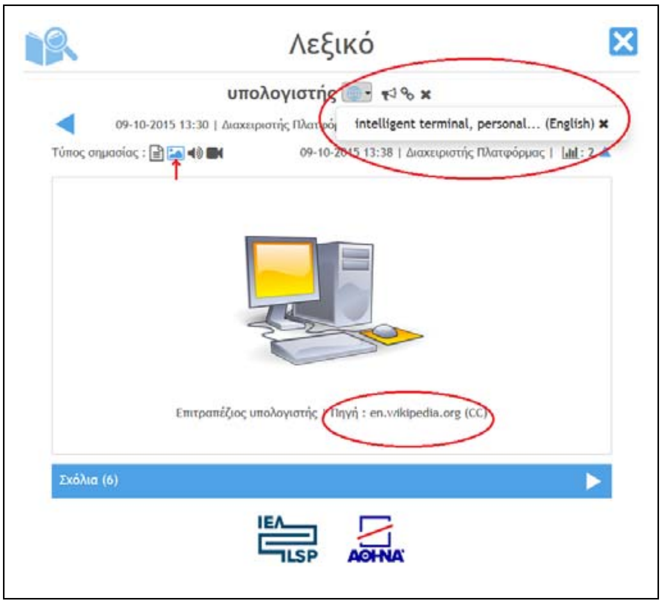


Figure 6: Visualization material associated with the term "computer" is selected.

3.2 Terminology Collection via Crowd Sourcing

Ad hoc terms are those terms translated into a receiver language but not yet undergone the full validation process. Most commonly, ad hoc terms are created within small user groups to serve specific communication needs within these groups. In many cases and for a considerable time phase, there may co-exist parallel versions of ad hoc translations of the same term used in the literature as they derive from different micro-environments, like, for example, different laboratories or university departments in the same national language environment.

This fact alone is a strong deviation from the definition of what a term is, where unique and unambiguous reference are major term properties (Wüster, 1979).

This phase corresponds to step 1 of the validation procedure referred to in section 3.1, and also is the environment from which most usually crowd sourced term collections derive. However, terms collected within this framework still may prove of great value, especially when the need for knowledge communication and understanding is critical as in the case of educational material for deaf student integration in a mainstream university or vocational training environment.

As regards the TEI-A platform, the crowd sourcing option has been a design prerequisite (Tedjamulia et al., 2005), (Wang et al., 2012) aiming to activate volunteers towards the target of collecting as many terms as possible from different sources, in order to facilitate deaf students' integration into the academic environment (Tausczik et al., 2014).

Thus, the platform accommodating the lemmas from the general purpose language sector along with the various terminology collections has been designed to be open to

crow sourcing activities (Doan et al., 2011) for enrichment with new lemma entries.
Following the main *Open eClass* pattern, three basic user roles are supported:

i. Student,
ii. Instructor, and
iii. Administrator.

Individuals from these three categories have equal access rights as regards viewing of educational material, but they also have scalable rights in respect to adding material to the platform. Moreover, the platform also supports special intermediary roles such as "*administrator assistant*", "*user administrator*", "*group leader*" and "*visitor*".
These roles serve, among other functionalities, the goal for lexical material enrichment through crowd sourcing (Sun et al. 2012).
Authorized users may enter new terminology items including the term definition and the whole scale of supporting multimedia material (icons, video, text etc), and they can modify or completely delete entries.
Non authorized users may equally propose new terminology items, being allowed to provide all types of material associated with a term. For security and quality control reasons, however, the items added by non authorized users do not become automatically visible to the whole user community.
Every new addition needs to undergo quality control for content and media (i.e. video quality) before becoming available to the platform user community. Thus, only authorized users, including domain experts from the GSL signers' community, can validate terms suggested by non authorized users and make them visible to the whole user community. It should be mentioned that in this case validation does not correspond to the ISO prescribed procedure, since there is no possibility to go through all term validation steps as foreseen. However, control is performed for content by native GSL signers who are members of the TEI-A community.

The overall terminology enrichment actions open to crowd sourcing initiatives incorporate:

1. New lemma or new sense entry
2. Modification of a lemma or a sense
3. Validation of a proposed lemma sense
4. Communication or hiding of a lemma sense or a lemma description
5. Linking of a lemma with a lemma in a different language (Greek and/or English)

Figure 7: Educational content accessibility support buttons available throughout navigation. Encircled the lexicon activation button.

All information gathered as a result of the above actions is easily accessed via the graphical user interface (GUI) within the lexicon window of the platform. The lexicon window may be activated at any point of the navigation process by clicking on the "book and lens" button (Fig. 7).

4. Conclusion

Here we presented an approach to supporting access of Deaf students to university level educational material via incorporation of GSL elements in a content presentation platform that incorporates elements which serve *Access for All* principles. GSL video presentations of various content parts are supplemented by a lexicon environment incorporating a bilingual general purpose dictionary for the language pair GSL-Greek, and a number of trilingual (GSL-Greek-English) terminology glossaries aimed to support written text understanding. For the enrichment of the terminology glossaries available in GSL, the platform has been implemented following an open design that targets to raising contributions via crowd sourcing.
Since terminology translation validation is a complicated time and human resources consuming enterprise, crowd sourced enrichment of existing resources seems to provide a solution to the severe problem of lack of educational aids in GSL, despite the discussion on ad hoc term status and the more general questioning of crowd sourcing effectiveness (Zhao & Zhu, 2014) after the first period of vast expansion of crowed sourced information collection.

5. Acknowledgements

We acknowledge support of this work by the project "Computational Science and Technologies: Data, Content and Interaction" (MIS 5002437) which is implemented under the Action "Reinforcement of the Research and Innovation Infrastructure", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).
P. Kakoulidis has been a member of ILSP SLT team during the reported platform development period.

6. Bibliographical References

Chamberlain, Ch., Morford, J., Mayberry, R. (1999). *Language Acquisition By Eye*. Psychology Press.
Dimou, A-L., Goulas, T., Efthimiou, E., Fotinea, S-E., Karioris, P., Pissaris, M., Korakakis, D., Vasilaki, K. (2014). Creation of a multipurpose sign language lexical resource: The GSL lexicon database. Proc. of *6th Workshop on the Representation and Processing of Sign Languages*, LREC 2014, Reykjavik, Iceland, pp. 37-42.
Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. Communications of the ACM, 54(4), 86-96.
Efthimiou, E., Fotinea, S-E., Goulas, T., Kakoulidis, P. (2015). User friendly Interfaces for Sign Retrieval and Sign Synthesis. In M. Antona, M. & C. Stephanidis (Eds.). Proc. of *9th International Conference, UAHCI 2015, Part II*, held as Part of HCI International 2015, Los Angeles, CA, USA, August

2-7, 2015, LNCS vol. 9176, pp. 351-361, Springer, Heidelberg.
Efthimiou, E. “Making Educational Content Accessible for the Deaf: The development of a multi-level platform”. In *Proceedings of the 2nd Colloquium on lesser user languages and computer linguistics*, pp: 181-194. Bolzano 13-14 Nov. 2008.
Efthimiou, E. & Fotinea, S-E. (2007). “An Environment for Deaf Accessibility to Educational Content” in *New Trends in Information and Communication Technology & Accessibility, Proc. of ICTA ’07*, April 12-14, Hammamet, Tunisia, pp.125-130.
Fotinea, S.-E., E. Efthimiou and A.-L. Dimou (2012). Sign Language computer-aided education: exploiting GSL resources and technologies for web deaf communication *Proceedings of ICCHP-2012 the 13th International Conference on Computers Helping People with Special Needs*, July 2012, Linz, Austria.
Fotinea, S.-E. and Efthimiou, E. (2009). A platform for sign language content presentation. *Proc. of the V International Conference on Multimedia, Information and Communication Technologies in Education* (m-ICTE 2009), April 09, Lisbon, Portugal, pp. 681-685.
Ghari, Z. (2016). The Cognitive, Psychological and Cultural Impact of Communication Barrier on Deaf Adults’ Content of Speech in Iran. *Commun Disord Deaf Stud Hearing Aids 4:164*. doi:10.4172/2375-4427.1000164.
Gruber, T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition 5/2*, pp:11-30.

ISO/R 704 (1968). *Naming Principles*. International Standard Organization.
ISO/R 1087 (1969). *Vocabulary of Terminology*. Geneva, International Standard Organization.
ISO 704 (1987). *Principles and methods of terminology*. Geneva, International Standard Organization.
ISO 1087 (1990). *Vocabulary of Terminology*. Geneva, International Standard Organization.
ISO 1087/1 (2000). *Terminology work – Vocabulary – Part 1: Theory and application*. Geneva, International Standard Organization.
Katsoyannou, M. & Efthimiou, E. (2004). Terminology theory, methodology and practice. In Katsogyannou, M. & Efthimiou, E. (eds) *Terminology in Greek: Research and Applications*. Kastaniotis, Athens, Greece, pp: 25-60 (in Greek language).
Klima, E., and Bellugi, U. (1979). The signs of language, Harvard University Press, USA, pp.205
Leigh, I. & Andrews, J. (2017). *Deaf People and Society: Psychological, Sociological and Educational Perspectives*. Routledge, New York. (2017, 2nd edition).
Liddell, S. and Johnson, R. (1986). American Sign Language Compound Formation Processes and Phonological Remnants, In *Natural Language and Linguistic Theory,* nol.4, Reideil Publishing Co, pp.445-513.
Phau, R., and Josep, Q., (2010), Nonmanuals: their grammatical and prosodic roles., Sign Languages, In D. Brentari (ed). 381-402. Cambridge: Cambridge University Press.
Sager, J.R. (1994). Terminology: Custodian of knowledge and means of knowledge transfer. *Terminology 1/1*, pp:7-16.
Sandler, W., Lillo – Martin, D. (2006). *Sign Language and Linguistic Universals*, Cambridge University Press, UK, pp.72.
Šumak, B., HeričKo, M., & PušNik, M. (2011). A meta-analysis of e-learning technology acceptance: The role of user types and e-learning technology types. *Computers in Human Behavior*, 27(6), 2067-2077.
Sun, Y., Fang, Y., & Lim, K. H. (2012). Understanding sustained participation in transactional virtual communities. *Decision Support Systems, 53(1)*, 12-22.
Tausczik, Y. R., Dabbish, L. A., & Kraut, R. E. (2014). Building loyalty to online communities through bond and identity-based attachment to sub-groups. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 146-157). ACM.
Tedjamulia, S. J., Dean, D. L., Olsen, D. R., & Albrecht, C. C. (2005, January). Motivating content contributions to online communities: Toward a more comprehensive theory. In *System Sciences*, 2005. HICSS’05. *Proceedings of the 38th Annual Hawaii International Conference on* (pp. 193b-193b). IEEE.
Wang, H., Chung, J. E., Park, N., McLaughlin, M. L., & Fulk, J. (2012). Understanding online community participation: A technology acceptance perspective. *Communication Research, 39(6)*, 781-801.
Wüster, E. (1979). *Einführung in die Allgemeine Terminologielehre und Terminologische Lexicographie*. (2 Volumes) Wien/New York, Technische Universität Wien.
Zhao, Y., & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*, 16(3), 417-434.

7. Language Resource References

DIOLKOS (DVD-ROM), 2006. ISBN 960-8191-39-4, http://www.ilsp.gr/diolkos.html

The POLYTROPON Parallel Corpus

Eleni Efthimiou, Kiki Vasilaki, Stavroula-Evita Fotinea, Anna Vacalopoulou,

Theodoros Goulas, Athanasia-Lida Dimou

ILSP / "Athena" R.C.
Artemidos 6 & Epidavrou, Maroussi, GR-15125, Greece
{eleni_e, kvasilaki, evita, avacalop, tgoulas, ndimou}@ilsp.gr

Abstract

Here we present the POLYTROPON parallel corpus for the language pair Greek Sign Language (GSL) – Modern Greek, which is created and annotated aiming to serve as a golden corpus available to the community of SL technologies for experimentation with various approaches to SL processing, focusing on machine learning for SL recognition, machine translation (MT) and information retrieval. The corpus volume incorporates 3,600 sentences performed by a single signer in three repetitions each, captured in front view by means of one HD and one kinect camera. Corpus creation was based on the validation procedure of a set of 2,000 lemmas deriving from the GSL segment of the Dicta-Sign corpus. Annotation of the corpus has provided interesting results in relation to all representation levels discussed within grammar theory, namely, lexicon, morphology, syntax, and semantics. Furthermore, it has allowed extraction of initial feature sets with the aim to reach a GSL level of abstraction close to the one currently available for Greek language representations, exploiting the inherent characteristics of the language. The POLYTROPON corpus is available to the SL research and SL technologies community via the CLARIN:EL infrastructure.

Keywords: SL data acquisition, GSL-Greek bilingual annotated resource, SL technologies, SL-Text parallel golden corpus, SL-based machine learning

1. Introduction

In the framework of research activities undertaken within the POLYTROPON project[1], significant effort was placed in maintaining and extending a Greek Sign Language lexicon dataset which consisted of lemmas captured by means of diverse capturing devices, lemma list construction methodologies and approaches for verification of acceptance by the local deaf community. The aim of this venture was to unify GSL lexical resources acquired during a time span of approximately fifteen years of different acquisition phases. The methodological principles and rationale for revisiting and recapturing the existing GSL lexicon resources have been reported in Dimou et al. (2014) based on the usability plan of the database designed to accommodate the new lexicon resource.

Thus, the POLYTROPON lexical database (POLYTROPON Bilingual Dictionary, 2015) was created with a threefold goal: i) to gather and recapture already available lexical resources of GSL in an up-to-date homogeneous manner, ii) to enrich these resources with new lemmas, and iii) to end up with a multipurpose-multiuse resource, which is equally exploitable in end user oriented educational/communication services but also in developing various SL technologies, including information extraction, Web accessibility tools, incorporation of lexical information in natural language processing (NLP) systems for SL processing as in the case of machine translation from and into sign language, creation of training material for sign recognition technologies and input to sign synthesis tools enabling signing by virtual signers (i.e. avatars).

In Efthimiou et al. (2016), a detailed account of the newly acquired lexicon resource provided information as regards the features associated with each lemma in the POLYTROPON lexicon database, as well as the way this information has been visualized (Fig. 1) to make the lexicon content accessible by end users outside the SL research community, mainly targeting: (i) the bilingual education of deaf children, and (ii) of the learning of GSL as a second language (L2).

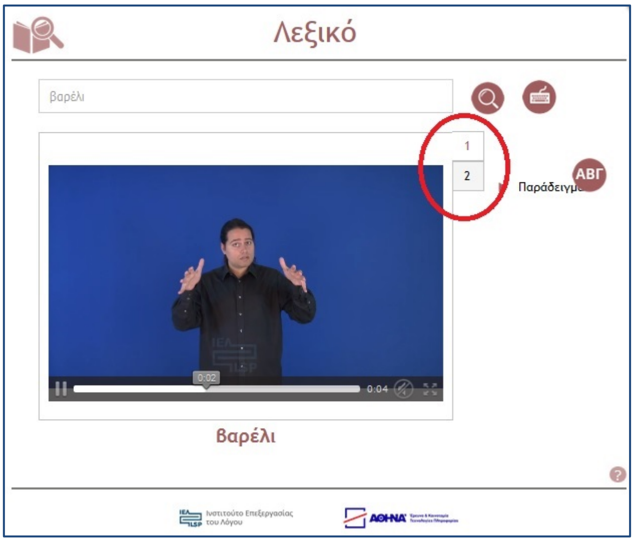


Figure. 1: Indicative snapshot of the visualization environment of the POLYTROPON lexical database. Here, two GSL synonyms are linked to one sense and one corresponding lemma in Modern Greek.

Efthimiou et al. (2016) provided examples of use of the POLYTROPON lexical resource in two educational platforms, namely, the official educational content platform for secondary education in Greece, and an

[1] http://www.ilsp.gr/el/infoprojects/meta?view=project&task=show&id=198

e-class platform as adapted by the Technical Vocational Institute of Athens (TEI-A), demonstrating the usability of this resource in the context of SL technologies based on lemma matching, such as dynamic synthetic signing and written text accessibility (Fig. 2).

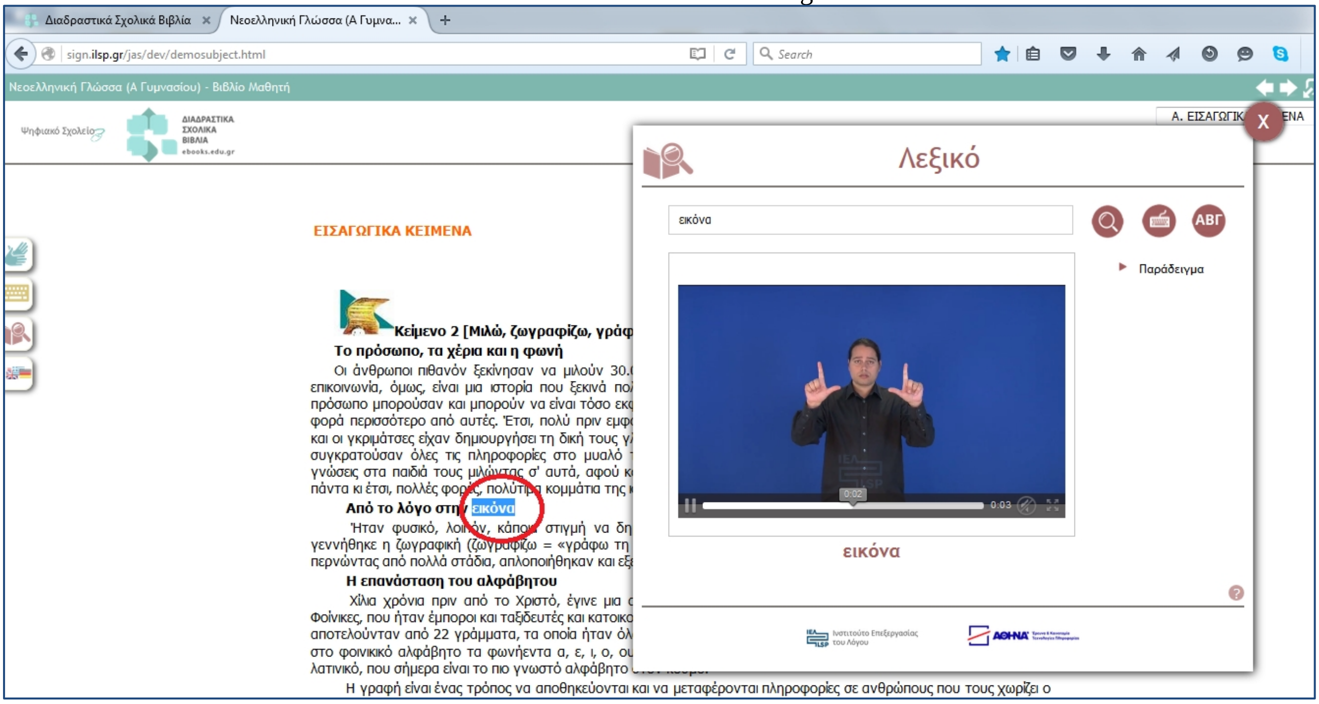


Figure 2: Web text accessibility tool exploiting GSL lexicon database content.

In the rest of the paper, we will refer to the specific segment of the POLYTROPON resource, which is composed of GSL sentences added as examples of use for lexicon lemmas and their Greek translations. This set of data formed an independent parallel corpus resource, extensively annotated to serve development of SL technologies that crucially rely on availability of a “golden” corpus for machine learning purposes.

2. Corpus Content and Acquisition Methodology

As previously mentioned, the main objective that led to the creation of the POLYTROPON corpus was to build a bilingual parallel corpus for the language pair Greek Sign Language – Modern Greek that could serve as a “golden” corpus available to the community of SL technologies for experimentation with various approaches to SL processing, focusing on machine learning for SL recognition, MTand information retrieval (Efthimiou et al., 2015).
Corpus creation was based on the validation procedure of a set of 2,000 lemmas, originally derived from the GSL segment of the Dicta-Sign corpus (Matthes et al., 2012) in a three-step process:

- *In step 1*, all lemmas were spontaneously commented on by a working group of experts during unofficially recorded sessions.
- *In step 2*, selected sentences from these discussions were recorded in three repetitions each, in studio conditions at a later stage. Among the working group there was consensus that the selected sentences also form good examples of use of the discussed lemmas, so that they are eligible for inclusion in the lexicon database in the “example of use” information column.
- *In step 3*, one out of the three repetitions of each recorded sentence was annotated in iLex (Hanke & Storz, 2008; Efthimiou et al., 2016).

Annotation of the sentences on gloss level revealed the use of new lemmas within sentence content, not initially included in the lexicon. Thus, the above described procedure of lemma validation was repeated for the new lemmas we well. Discussion of new lemmas resulted in new sentences in an iterative process, enriching the GSL iLex lexicon DB with 1,600 new lemmas in total, while commentaries on new entries generated the new clauses which completed acquisition of the content of the POLYTROPON corpus following baseline elicitation principles as in Matthes et al. (2010).
In total, the POLYTROPON parallel corpus incorporates 3,600+ clauses in three repetitions each, captured in front view by means of one HD and one kinect camera.
In the next section, an account of the adapted annotation scheme and annotation findings is provided.

3. Corpus Annotation Scheme and Annotation Findings

3.1 Corpus Annotation Scheme

The corpus Annotation scheme (Fig. 3) entails the following tier set:

"*Clause*" defining clause boundaries of the signed utterances,
"*Gloss*" assigned to each identified token and provided in Greek,
"*Greek equivalent clause*", which provides the Greek translation of each signed unit within the "Clause" time frame,
"*Classifier*", which provides one tier for each classifier handshape and allows feature values assignment with respect to one- or two-hands and same or different activity and the classifier's semantic content (Fig. 4), using the labels [entity], [shape], [handling] and [predicative],
"*S-type*" to mark main vs. subordinate constructions,
"*S-category*" for the marking of sentence categories according to syntactic classification that receives feature values based on classical descriptive grammar classification.

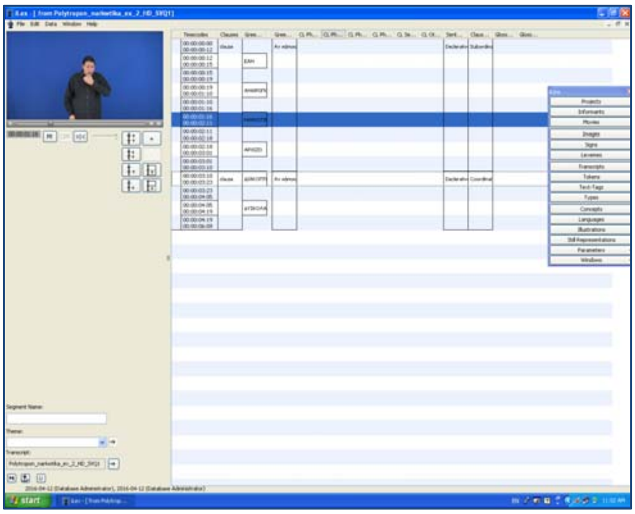


Figure 3: The overall POLYTROPON corpus annotation scheme in iLex.

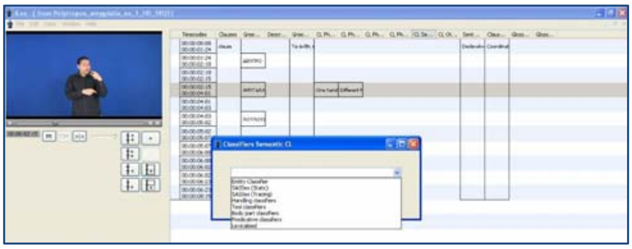


Figure 4: Classifier annotation according to semantic function.

Annotation was performed by a coda GSL expert in the iLex annotation environment and was cross-checked by two more linguists with expertise in annotation and analysis of GSL data.
The translation of the annotated sentences in Greek was performed in two phases:

a) a strongly GSL-influenced initial translation version provided directly by the annotator, followed by
b) a "corrected" translation version, which provided fully acceptable Greek sentences as regards naturalness and grammaticality, performed by an expert in Greek language.

The overall annotation scheme was designed to provide a range of information expressed in terms of lexical and sentential feature bundles, aiming to allow for search options targeting morphology, semantics and syntax relevant events (Liddell and Johnson, 1986; Pfau and Quer, 2010; Quer et al., 2017).
Regarding representation-level related information, it must be noticed that although no articulation (phonological-phonetic) information is visible in the corpus as regards sign tokens, this is directly available in the POLYTROPON lexical database (Efthimiou et al., 2016), where lemmas receive full phonetic articulation descriptions according to the HamNoSys notation system (Prillwitz et al., 1989; Hanke, 2004) and the SiS-Builder embedded non-manuals notation tool (Goulas et al., 2010; Efthimiou et al., 2014).
By the completion of the annotation process, the POLYTROPON corpus proved to be a rich source of GSL grammar information, equally useful to SL technologies oriented research and SL theoretical linguistic analysis.

3.2 Annotation Findings

In this section, the major linguistic findings of the annotation process are listed, since they bring insight as to a series of phenomena SL technologies need to tackle.

Gloss-level findings
Gloss-level annotation has made visible three types of lemmas not previously contained either in the iLex lexicon or in the GSL-Greek bilingual lexicon database in the SiS-Builder environment. These involve:

a) A set of new lemmas which were not previously included in any of our two databases. These were directly added in the iLex lexicon during annotation process, and also created a set of new reference type entries which enriched the SiS-Builder lexicon database. This procedure enriched both databases with 1600 entries in total.

Figure 5: Sequence in manual and non manual activity when articulating the GSL expression EARS-DOWN.

b) A set of GSL-specific expressions with no direct translation equivalent in Greek or English, such as EMPTY-POCKETS to imply the meaning of "I am broke" or EARS-DOWN to express the meaning of "obey" (Fig. 5). These were classified under "GSL special expressions" in both lexicon databases.

c) A set of exclamation gestures with semantic value in direct equivalence to embodied signals of oral expression as, for instance, the embodied exclamatory expression adding affect-related extra-linguistic information to strengthen utterances such as “what can I say!”, “I don’t know!” etc. (Fig. 6).

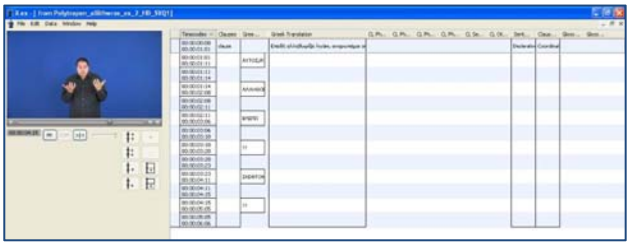


Figure 6: Embodied extra-linguistic expression commonly met in GSL and Greek.

Compounds
Regarding compound formations, the POLYTROPON corpus annotation allowed for identification of the following compounding options in line with formal descriptions as in Sandler & Martin (2006):

- *classifier+classifier*: as in formation of the GSL sign “lighthouse” incorporating the scheme: [CL-5C+CL-5]
- *sign+sign*: as in formation of the GSL sign “air hostess”: involving the sign combination [AIRPLANE]+[ACCOMPANY], or the sign “cow” exploiting [ANIMAL]+[MILK].
- *sign+ classifier*: as in formation of the GSL sign “pilot”, exploiting the combination: [AIRPLANE]+[CL-S]
- *classifier+sign*: as in formation of the GSL sign “letter”, by means of the combination [CL-C1]+[SEAL]
- *ad-hoc formations*: such formations involve concatenation of signs, as for example in [PAPER-NOTE-REMEMBER] to express the meaning of “take notes”, and are characterized by their unique appearance in the entire corpus. This set of lemmas has not been incorporated in the lexicon databases yet, since they need to be further cross-checked with more native signers in order for their compound status to be validated.

Classifiers
As regards the classifier content of POLYTROPON, annotation was performed (i) on morphophonemic level involving markings related to formation and including information as to handshape, two- or one-hand activity, and same or different activity performed by the hands, and (ii) on semantic-content level, assigning the labels: [entity] [shape] [handling] and [predicative] from a drop-down menu, following an analysis as in Efthimiou et al. (2010).

Sentence-level findings
On sentence level, two major clause categories were annotated, that is, *main* and *subordinate*.
Under main clauses, there are also coordination constructions which in many cases replace oral language subordination as when coordination makes use of the INDEX/TOPIC mechanism, which replaces Relative Clause subordination, met in a number of oral languages.
Subordinate clause formation involves constructions which present clear subordination markers extensively indicated via the combination of manual and non-manual signals as in the case of the causative marker [BECAUSE] (Fig. 7), or in conditional constructions where the semantics of “if” are expressed both via manual and non-manual elements and where presence of the non manuals is obligatory.

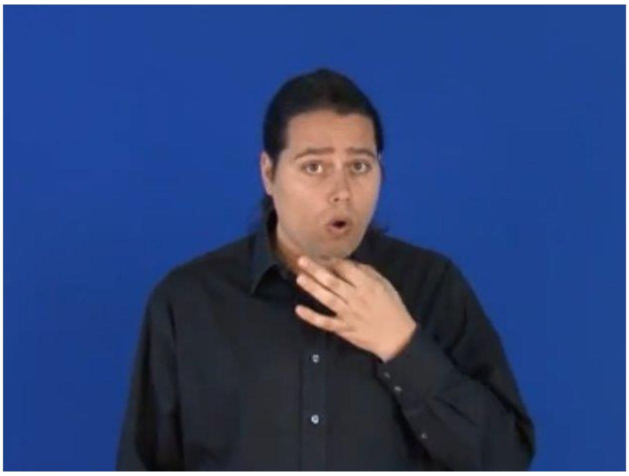


Figure 7: The causative marker [BECAUSE] that introduces subordinate Causative Clause in GSL.

Sentence type
For both main and subordinate clauses, further classification assigns sentence category values from the following list:

- Declarative-Affirmative
- Declarative-Negative
- Interrogative (Yes/No, Wh)
- Rhetoric Q&A
- Imperative
- Exclamation

Rhetoric Q&As are declarative constructions, which incorporate a WH-like utterance without exhibiting the full non-manual activity usually present in WH-questions, aiming to enforce the focus catching effect of the signed message. Examples of rhetoric Q&A are utterances like [TODAY-EAT-WHAT-RICE-WITH-CHICKEN], which expresses the message “Guess what I will eat today! Rice with chicken!”.

3.3 Corpus Exploitation

It has already been mentioned that creation of the POLYTROPON parallel corpus has been directed towards its application in the SL technologies domain, mainly targeting the need for annotated data. Given work with the language pair GSL-Greek, the main aim has been to provide a significant amount of GSL data annotated, which may allow reaching a similar level of abstraction for both language representations. This abstraction should be succeeded by making exploitable the inherent characteristics of both languages, thus, reaching a state where we can apply deep learning experiments on GSL data, where representation of both words and signs takes the form of a vector of characteristics as in (Fig. 8).

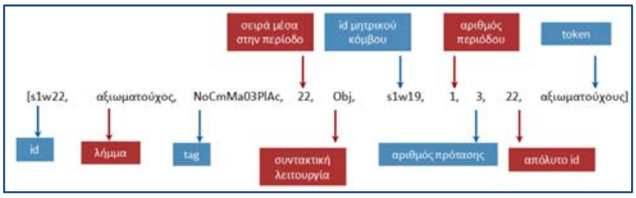


Figure 8: Feature vector representation of lexical items.

Furthermore, since the goal of the acquisition team has been to provide the research community with a golden corpus for machine learning in the areas of SL corpus mining and MT, the current release of the POLYTROPON parallel corpus is available via the *clarin:el* repository, which is the Greek sector of CLARIN[2], the European infrastructure for language resources and technology. The corpus is available free of charge but subject to Creative Commons (*CC - BY*) licensing[3]. For its identification in *clarin:el*, the POLYTROPON parallel corpus has been assigned the persistent identifier (PID):

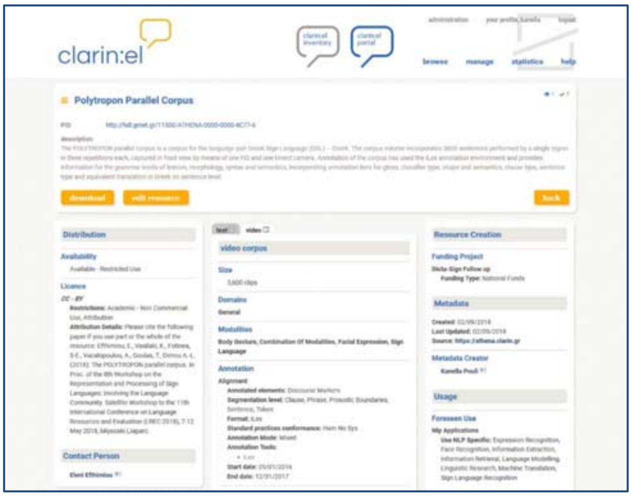


Figure 9: Word/sign representation as a vector of characteristics.

http://hdl.grnet.gr/11500/ATHENA-0000-0000-4C77-6 (Fig. 9), while academic users may directly reach the resource in the *clarin:el* platform by following the link: https://athena.clarin.gr/resources/browse/polytropon-parallel-corpus/197061c20d9711e89c26aa3fc8d33ad8b716f4f795884a8792b708207d02bd84/.

[2] CLARIN: Common Language Resources and Technology Infrastructure (www.clarin.eu).
[3] https://creativecommons.org.

Regarding sentence-level representation, experimentation is currently oriented towards exploiting Dependency Tree Structure representations of input text and signed clauses using Tree Editor TrEd 2.0.

4. Conclusion and Future Plans

The POLYTROPON parallel corpus was created to mainly address SL processing needs in the framework of human language technologies applications seeking mainly ways to extend our current knowledge with respect to corpus-based and statistical approaches to MT, but also in service of SL technologies with focus on sign recognition, information extraction and information retrieval from video sources. The resource aims to trigger new challenges both on technological and SL linguistic grounds. In this context and in order to better serve the goal of exploiting the corpus in the context of machine learning by providing a multi-signer approach to the already acquired data, a crowd-sourcing activity is currently planned, which will invite native GSL signers to repeat a selected segment of the corpus content with the aim to enrich the variability of signers and signing space conditions towards serving machine-learning purposes more effectively.

5. Acknowledgements

We acknowledge support of this work by the project “Computational Science and Technologies: Data, Content and Interaction” (MIS 5002437) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

6. Bibliographical References

Dimou, A-L., Goulas, T., Efthimiou, E., Fotinea, S-E., Karioris, P., Pissaris, M., Korakakis, D., and Vasilaki, K. (2014). Creation of a multipurpose sign language lexical resource: The GSL lexicon database. Proceedings of 6th Workshop on the Representation and Processing of Sign Languages, LREC’14, Reykjavik, Iceland, pp. 37–42.

E. Efthimiou, S-E. Fotinea, A-L. Dimou, T. Goulas, P. Karioris, K. Vasilaki, A. Vacalopoulou, M. Pissaris, and D. Korakakis. (2016). From a sign lexical database to an SL golden corpus – the POLYTROPON SL resource. Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining. Satellite Workshop to the 10th International Conference on Language Resources and Evaluation (LREC-’16), 28 May 2016, pp. 63–68, Portoroz, Slovenia.

Efthimiou, E., Fotinea, S-E., Goulas, T., and Kakoulidis, P. (2015). User friendly Interfaces for Sign Retrieval and Sign Synthesis. In M. Antona, M. & C. Stephanidis (Eds.). Proceedings of 9th International Conference, UAHCI 2015, Part II, held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, LNCS vol. 9176, pp. 351–361, Springer, Heidelberg.

Efthimiou, E., Dimou, A-L., Fotinea, S-E., Goulas, T., and Pissaris, M. (2014). SiS-builder: A Tool to Support Sign

Synthesis. Proceedings of 2nd Int'l Conference on the Use of New Technologies for Inclusive Learning, pp. 26–36. York, UK.

Efthimiou, E., Fotinea, S-E., and Dimou, A-L. (2010). Towards decoding Classifier function in GSL. In Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010), 2010, Satellite workshop of the LREC-'10 Conference, Valetta, Malta, pp.76–79.

Goulas, T., Fotinea, S-E., Efthimiou, E., and Pissaris, M. (2010). SiS-Builder: A Sign Synthesis Support Tool. In Dreuw, P. et al. (eds.), Proceedings of 4th Workshop on Representation and Processing of Sign Languages, LREC-'10, pp. 102–105.

Hanke,T. (2004). HamNoSys - representing sign language data in language resources and language processing contexts. Proceedings of 1st Workshop on Representing and Processing of Sign Languages, LREC-'04, pp. 1–6.

Hanke, T. and Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. *Construction and Exploitation of Sign Language Corpora.* Proceedings of 3rd Workshop on the Representation and Processing of Sign Languages. ELRA, Paris, 64–67.

Klima, E., and Bellugi, U. (1979). The signs of language, Harvard University Press, USA, pp.205

Liddell, S. and Johnson, R. (1986). American Sign Language Compound Formation Processes and Phonological Remnants, In *Natural Language and Linguistic Theory,* vol.4, Reideil Publishing Co, pp.445–513.

Matthes S., Hanke T., Regen A., Storz J., Worseck S., Efthimiou E., Dimou A.-L., Braffort A., Glauert J., and Safar. E. (2012). Dicta-Sign – Building a Multilingual Sign Language Corpus. Proceedings of 5th Workshop on the Representation and Processing of Sign Languages, LREC-'12, Istanbul, Turkey.

Matthes S., Hanke T., Storz J., Efthimiou E., Dimou A-L, Karioris P., Braffort A., Choisier A., Pelhate J., and Safar E. (2010). Elicitation tasks and materials designed for Dicta-Sign's multi-lingual corpus, LREC '10, Valetta, Malta.

Phau, R. and Josep, Q., (2010). Nonmanuals: their grammatical and prosodic roles, Sign Languages, In D. Brentari (ed). 381–402. Cambridge: Cambridge University Press.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T., and Henning, J. (1989). HamNoSys. Version 2.0. *Hamburg Notation System for Sign Language: An Introductory Guide*. Signum Verlag, Hamburg.

Quer, J., Cecchetto, C., Donati, C., Geraci, C., Kelepir, M., Pfau, R., and Steinbach, M. (eds): (2017). *SignGram Blueprint: A Guide to Sign Language Grammar Writing*. DE GRUYTER MOUTON. ISBN 978-1-5015-1180-6.

Sandler, W. and Lillo-Martin, D. (2006). Sign Language and Linguistic Universals, Cambridge University Press, UK, pp.72.

7. Language Resource References

POLYTROPON Bilingual Dictionary. (2015). Clarin:el repository persistent identifier (PID): http://hdl.grnet.gr/11500/ATHENA-0000-0000-42D5-5.

POLYTROPON Parallel Corpus. (2017). Clarin:el repository persistent identifier (PID): http://hdl.grnet.gr/11500/ATHENA-0000-0000-4C77-6

Extending the AZee-Paula Shortcuts to Enable Natural Proform Synthesis

Michael Filhol[1,2], John McDonald[3]
[1] LIMSI (CNRS), [2]Université Paris–Saclay, [3]DePaul University
[1] Univ. Paris–sud, bât. 508, 91400 Orsay, France
[3] 243 S. Wabash, Chicago, IL, USA
michael.filhol@limsi.fr, [3]jmcdonald@cdm.depaul.edu

Abstract

Proform structures such as classifier predicates have traditionally challenged Sign Language (SL) synthesis systems, particularly in respect to the production of smooth natural motion. To address this issue a synthesizer must necessarily leverage a structured linguistic model for such constructs to specify the linguistic constraints, and also an animation system that is able to provide natural avatar motion within the confines of those constraints. The proposed system bridges two existing technologies, taking advantage of the ability of AZee to encode both the form and functional linguistic aspects of the proform movements and on the Paula avatar system to provide convincing human motion. The system extends a previous principle that more natural motion arises from leveraging knowledge of larger structures in the linguistic description.

Keywords: Sign Language Linguistics, Proforms, Classifiers, Synthesis, Avatars

1. Introduction

Producing natural synthesis of sign languages using an avatar is a goal with far-reaching applications for Deaf-hearing communication including improving sign language education tools, anonymizing online communication in sign language, and enabling translation for situations where hiring a certified interpreter is impossible. To support all of these applications, a sign synthesizer must be able to express all aspects of sign language including the full range of body signals used to communicate.

Sign languages use a range of linguistic processes to communicate, the most basic of which being those listed in sign dictionaries. These are gestural units that have a standardized and stable meaning–form association. In addition, signers use a range of grammatical processes for rich, natural communication. All of these processes are communicated through signals involving the arms, eyes, face, torso and neck of the signer. Natural sign synthesis remains a challenge partially because of the fact that these structures can overlap and interact on the body (Weast, 2011) and animating such structures requires leveraging both sign language linguistics and knowledge of human motion (Braffort et al., 2015).

Most current sign synthesis systems are able to animate a stream of lexical signs (Wolfe et al., 2011; Elliott et al., 2008; Lombardo et al., 2011). However, more freehand constructs in sign languages such as classifier placement and movement or size and shape specifiers remain a challenge because of the inherent variability in form expressed through the signer's body. These structures use configurations on the signer's body known as proforms, wherein a part of the signer's body stands in for an object and is often iconic of its shape. For example, in American Sign Language (ASL) and French Sign Language (LSF), the index finger oriented vertically will represent a standing person, whereas a "C" handshape will represent a cylindrical object such as a glass (Liddell, 2003). Such proforms can be used to express the placement or movement of objects or to describe their size and shape.

Proform structures have traditionally been a challenge for synthesis systems, both from a linguistic and from an animation standpoint, because they are highly productive. The placement and movement of body articulators cannot easily be captured by pre-set configurations, and so cannot be pre-animated or recorded. Thus systems relying on motion capture (Gibet et al., 2011) or reusable hand animation (Wolfe et al., 2011) must fall back on more primitive synthesis techniques. In addition, while many proforms are predefined for a given sign language, variations in the shape or configuration of an object can be expressed through a near infinite range of hand or body configurations, especially when one considers size and shape specifiers (Liddell, 2003).

Prior efforts have relied either on synthesis from phonetic components (López-Colino and Colás, 2011) or on predefined templates that encapsulate a limited set of standard proforms (Huenerfauth, 2004). In both cases the results were far from natural involving awkward body configurations and robotic motion. The present work addresses three weaknesses that contribute to the robotic nature of the resulting animations.

1. Body gestures are never fully specified by the linguistics since there are a range of body configurations that can satisfy given linguistic constraints (Filhol and Braffort, 2006b). Synthesis systems have had to fill in the missing body constraints for example by overspecifying the motion linguistically.

2. Synthesized motion has been limited to the avatar's arms, whereas natural human arm motion is always accompanied by supporting torso and clavicle motion. (McDonald et al., 2016a).

3. Motion specifications were limited to key positions for the handshape and didn't specify dynamic differences such as acceleration or speed profiles that can profoundly affect perception.

To fully specify proform movement and placement, and to produce natural motion, the synthesizer must take into consideration both:

- the linguistically defined constraints, which abstract human motion into meaningful gestural units;
- the range of human motion that accompanies such gestural units, but which are not encoded linguistically.

This paper extends the work in (Filhol et al., 2017) which sought to bridge between a structured linguistic model of sign and a multitrack sign animation system. The system described here achieves natural linguistically driven proform motion through two key features. First, by separating the task into separate linguistic and animation components, it allows the linguistic component to encode the necessary information for the proform while allowing it to remain underspecified at a geometric level. This gives the avatar the needed freedom to move naturally within linguistic constraints. Second, it builds on the prior model's principle that natural motion is best achieved with large linguistic structures rather than from very basic phonetic specifications. The next section explores the perspectives of each system on proforms and what each offers for building a combined synthesizer.

2. Perspectives of the two systems on proforms

2.1. AZee for descriptions

The Sign Language description model AZee (Filhol et al., 2014) has several advantages that are relevant to the use of proforms in SL. The major one is its fully embedded geometric system that allows to build and describe points, vectors and paths in the signing space. From the origin of its predecessor Zebedee, geometric specification of body locations and skeletal orientations as points and vectors in an affine 3D space[1] have been an essential feature of the descriptions (Filhol and Braffort, 2006a). More than an alternative style of body posture description, such geometric approach accounts for at least three notable features of SL, which are difficult to capture with other, e.g. parametric, models.

First, it does not rely on a discrete set of points for locations in space, or directions for orientations. It allows to define geometric objects in a continuous space. In other words given two points, it is always possible to take the midpoint of the two. When making free productive uses of space, e.g. placing proforms to indicate relative positions, it is therefore possible to account for any relative placement, such as a date "in the middle of" the two boundaries of a delimited period on a time axis.

Second, dependencies between elements of the descriptions are made relevant. This is useful for depicting structures such as the proform placements or movements we are addressing because positions are often relative to (dependant on) each other. For example, in the predicate "rabbit near and on the right of tree", the target location point for the rabbit proform is relative to that of the tree. AZee allows to express the rabbit's position as a geometric translation of the point with the appropriate distance, instead of projecting to a grid defining everything relative to the chest.

Third, not only hands can be specified target locations but any articulator of the body. We have argued the necessity of this in several papers, but it becomes all the more relevant in dealing with depicting structures. Placing two-hand classifiers (e.g. round plates) and placing a full-arm classifier (e.g. tree) pose a problem if we want to consider hand placement alone as changing the classifier inside any expression will require a change of location as well. In AZee, a plate and a tree can be placed at the same location using the same point, even though hands actually end up in completely different locations.

Essentially, AZee is a language to write *rules* mapping invariant and parameterized *forms* to identified semantic *functions*, regardless of the level of linguistic description. Classifier/proform placements are indeed units that are difficult to locate in those terms, as they are arguably both lexical and grammatical constructions, or neither (Johnston categorizes those separately as "partly lexical signs"). AZee bypasses this problem, as any function-to-form link is tackled using the same description model. Like any other rule, a classifier generates a set of articulatory constraints for a consistently interpreted meaning.

For example, the upright index finger shape denotes a standing person, possibly with a wrist orientation depicting the direction in which the person is facing. Because the meaning conveyed with this finger arrangement is consistent, an AZee rule "proform-standing-person" can be defined to specify the appropriate articulatory constraints: index up, possibly facing along a parameterized direction, other fingers closed. As we stated with the AZee approach, only the set of necessary and sufficient constraints are to be specified. In this instance we therefore exclude palm orientation from "proform-standing-person", because only fingers matter. In contrast, a rule for "tree" would have to include all bones from the tip of the fingers down to the elbow, since not extending and spreading the fingers would result in breaking the meaning.

The AZee approach also encourages and facilitates factoring similar forms into new rules when the interpreted meanings share a common factor across multiple productions. For example, take the placements of entities in the signing space of the kind Liddell glosses with -BE-AT$^{\downarrow}$ suffixes. All are produced with a small settling movement towards the surface on which the object is placed (often downwards), and an eye gaze towards the target location. This is true regardless of what the object is, and regardless of what articulator set is conveying the object. In AZee, one would therefore factor this common form, parameterizing the proform *prf* and the location point *loc* to define a rule with semantic function:

"placement of *prf* in space at location *loc*"

producing the form:

[1] In geometry, an affine space is a vector space with no chosen metric or origin. In our case everything is defined relative to the body, including directions and distances. This allows implementation with any avatar. No body geometry is assumed by AZee.

"small straight movement of *prf* down to point *loc* + look at *loc* + synchronize eye gaze with a negative time offset".

Proform *prf* and point *loc* become empty placeholders, and necessary arguments of the created rule.
More factoring can take place, this time with a rule already found and reported elsewhere: the "category" rule, whose semantic function is to give a category in which to interpret a second argument item. It allows to juxtapose "town" and "Berlin", or "profession" and "bakery" to specify a sort of hypernym for the second item, likely but not necessarily ambiguous on its own. Classifier constructions in LSF often involve juxtapositions of a dictionary sign and a placement like the one described above, applied to a proform. The overall form then, including the juxtaposition, a slight head tilt and a specific inter-sign transition timing, is the same as that specified by the "category" rule. Plus, the meaning is to us all but similar: the first item gives a class of which the second is an instance.
Therefore, from the three rules below:

- category
- proform-vehicle
- place-prf

and a point *loc*, one can build the complex expression:

(E1) category(car(), place-prf(proform-vehicle, *loc*))

producing the sequence traditionally glossed as:

CAR VEHICLE-BE-AT$^{\downarrow loc}$

and taking care of the precise timing and adding the gaze towards *loc*.

2.2. Paula for Natural Animations of Sign

The Paula sign synthesis system compliments AZee's linguistic proposition in supporting natural animations of SL. As described in previous publications, it is a hybrid animation system that supports layering motion from a variety of sources including procedural, keyframe, etc. (McDonald et al., 2017). From the synthesizer's perspective, proform movements can be modeled as a collection of keyframe data. In this respect, Paula offers a range of features that allow it to produce more natural animation from such data, and allows leveraging animator/sign-expert skills to a larger degree than prior systems. The following features are key to producing natural animations of proform movement from the linguistic specifications:

1. Key postures can be set by using either forward (FK) or inverse-kinematics (IK) systems tuned for sign linguistics (McDonald et al.,). The IK system allows any point defined relative to the hand or arm to be placed at a chosen target in space, or at a chosen site on the avatar's body, and allows full exploration of the redundant degrees of freedom in the IK chain.

2. Keyframes can be scheduled completely asynchronously on different articulatory chains (McDonald et al., 2017).

3. Interpolations between key postures are accomplished using nonlinear rotation controllers that create transitions following the natural arcs of human motion. These interpolators also allow independent control of speed and trajectory along paths.

4. Procedural techniques automatically allow the torso and shoulders of the system to naturally support and accompany arm movements (McDonald et al., 2016a), and also add sub-linguistic ambient motion (McDonald et al., 2016b).

These features all contribute to allow the synthesizer to produce smooth, natural movement from the linguistic features described above.
The first of these features allows any articulator on or near the arm to be used for targeting, and facilitate the natural positioning of two-handed proforms such as in the *round-plate* classifier example from the last section. The fourth feature in this list extends the avatar's arm motion through the trunk all the way to the hips, directly addressing the second cause of robotic motion cited in the introduction. Finally and Paula's nonlinear rotation controllers provide for independent velocity control on articulatory chains, needed to animate the linguistic categories of dynamics, thus addressing the third cause of robotic motion.
To address the underspecified nature of proforms, Paula also offers components that help leverage animator and sign-expert skills as far, and as deep, as possible in the synthesis process. The first component most often used in sign animation is Paula's Sign Transcriber, which scripts for prerecordable segments (Wolfe et al., 2011). In spite of the fact that proforms are highly variable, Paula's Sign Transcriber allows animators to provide significant body posture and movement hints, which the proform generation system can then use to produce more natural animations than would be possible from the linguistic data alone. What remains is to build a coherent bridge between these two systems so that the necessary constraints are communicated while giving the animation system the freedom it needs.

3. Extending "the coarser the better"

3.1. Last proposal (and clarification)

The present work builds on the AZee-Paula bridge proposed in (Filhol et al., 2017), which was based on the principle that working with larger blocks of animation or procedural motion will generate more natural sign synthesis compared to animating from individual constraints and joint settings. The previous bridge mapped from AZee expressions that could be recognized when reading block descriptions output by the AZee parser. This allowed the animation system to shortcut the application of a block if a prerecorded or procedural animation is directly available, rather than reconstructing the entire block's animation from low-level primitives nested in the generated form description.
This principle, which the authors called "the coarser the better", relies on AZee's organization of scores as a hierarchical nesting of blocks, and asserts that the Paula animation system is able to produce more natural animation with larger parent blocks compared to combinations of many

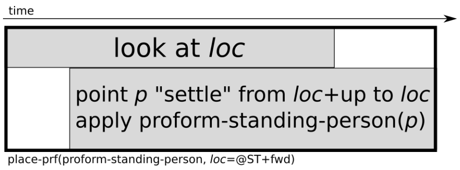


Figure 1: Score for expression (E2).

small child blocks. The system thus considers parent blocks in the hierarchy first, and falls back on developing child blocks only if no match is found.
The system accomplished this with a library of pre-animated segments and procedural techniques that correspond to blocks named after their AZee source expression. If the system recognized the expression in its entirety, the resulting animation or procedure would be invoked on the required portions of the avatar. In the absence of a match for an animation block, the system recursively falls back on animating child blocks until individual articulatory constraints would be required, sacrificing naturalness in the resulting synthesis. The multilinear nature of Paula's skeleton integrated these animation blocks into a seamless whole even when overlapping blocks simultaneously used a given set of articulations.

3.2. Limitation for Synthesizing Proforms

What the prior work did not insist on (and indeed possibly confused by basing its narrative mostly on functional AZee trees) is that all short-cuts described and exemplified were taken based on the nesting of the score blocks, therefore only dealt with the specified *forms* to animate. The contents of blocks in the shown XML structure contain a source AZee expression which is functional, but it was used as a mere string label for a stored form with which to short-cut. This means that nothing on the functional (semantic) side of the AZee system was considered during matching. This becomes a significant issue when considering proform synthesis.
For instance, consider applying this matching scheme to the proform expression below:

(E2) place-prf(proform-standing-person,
translate(@ST, along(medium, fwd)))

The previous system would first run the linguistic AZee interpreter to generate an XML specification of the resulting score of nested blocks (Filhol et al., 2017), illustrated in Figure 1. It would then search for a procedural or prerecorded animation short-cut whose name perfectly matched the source expression (E2), labelling the outer-most level of block nesting. It is unlikely that a perfect match, including the exact expression for the second argument, would be available as a prerecorded animation due to the infinite number of possible values for this argument. The system would then fall back on the block's child constituents (the inner blocks), which we see already consist of low-level constraints. Synthesizing from those would already be sacrificing naturalness. In this case such fallback is therefore a leap from too high a level to shot-cut, to one that is too low.

Yet the placement of a standing person does have a consistent natural dynamics, regardless of its location. There is a consistent handshape, with the upright index finger being slightly over-extended, and a consistent arcing motion, resulting from elbow and shoulder rotations which the description would specify as "downward" though it never comes out as such. If the animator has provided examples of this kind of motion, or if the system has a procedural specification for the right dynamics based on corps study, the animation system should be able exploit them, and producing in a more natural animation than one from a sparse "straight movement down" description. The "coarser the better" principle should somehow apply here despite the fact that the top-level expression cannot fully be matched.
To do this, the system needs a way to look into the AZee expression and enable some form of short-cutting there. In other words, in addition to matching blocks in the *form* tree as covered by our previous bridge, we need to expand our short-cutting scheme to look into the information on the *functional* side of the input as well, i.e. the AZee expression itself.
For example, to apply a generic proform placement procedure when animating (E2), we need to recognize part of its contents, with a template like the one below:

place-prf(proform-standing-person, X)

where X can be anything, provided it is an expression that evaluates to a point where to place the proform, and can be retrieved for actual use with the matched animation procedure.

3.3. Proposal for new system and results

To extend the "coarser the better + fall-back" principle, the new model proposes to allow an intermediate check for matching such templates via the following extended fall-back mechanism:

1. match as label for *form* shortcut;
2. match with template for *functional* shortcut;
3. recursively process child blocks if no match (recursion terminates when reaching a block consisting of low-level constraints only).

The matching in step 2 is similar to the characterization of classifier motions as abstract templates (Liddell, 2003), with a set of parameters provided by the linguistic system. It is then up to the animation system to read and interpret those parameters. For example, consider processing the expression below, where "midssp" is a pre-defined name representing a point in the centre of the signing space.

place-prf(proform-vehicle, midssp)

Assuming that Paula cannot find a full match for the complete expression, the animation system will perform the following steps. Notice that we are using the word template here both for the AZee expression matching and also for the animation data specified by the artist for the specific proform *proform-vehicle*.

1. look for a template procedure for *place-prf*;
2. look for a procedure or animated version of the ”proform-vehicle”;
3. Evaluate the *midsp* expression to obtain the point for placement;
4. Animate the avatar using the template data.

In evaluating the *proform-vehicle* argument, the system will look for an artist specified (or motion-capture if available) template to use for the animation. Paula will then be able to leverage a range of information from that template to use in the formation of the key-frames for the action. Among these will be the handshape, the torso and clavicle parameters, the arm and elbow height as well as the articulator point on or near the arm, used for targeting the proform.
In the case of the place-prf action above, the system would then set up two key-frames with associated velocity controls to “settle” the proform at point “midssp”. This settle action is an example of one of Johnson’s *ballistic* transitions that passes through the first position with a smooth speed and then eases to rest in the position of the second key. In setting up the keyframes and animating the segment, Paula will leverage the features described in section 2.2.. In particular, it will:

- read the handshape, initial elbow configuration, preferred comfortable height and other data for the proform from the artist template;
- use the IK system to set up the two keyframes for the motion (the first one with a target point that is above the final point by a small amount), adjusting the data from the artist template as needed;
- use the spine assist and livening procedures to move the avatar’s torso in concert.

From these keyframes the nonlinear motion controllers will move the avatar’s arms along natural arcs, which will be straight enough here to provide a perceptual ”straight-down” motion.
Note that in the first item above, the system reads a name for the handshape and triggers a procedure without looking into the articulatory constraints that compose it on the child level. The second item similarly applies a generic two-keyframe layout to implement a recognized AZee pattern. We emphasize that by doing so, the system is performing a shortcut on an element of the functional expression.
If at some point in this process Paula fails to find a match, for example if Paula does not have a template for the proform action, or the profrom specification deviates from one of the “known” forms, the system then falls back to the expression’s child blocks which will give a set of primitive constraints for the movement. Again, this would necessarily sacrifice quality but provides robustness for the system. This part of the system has not yet been integrated.
In addition to the *place-prf* template, the system currently also supports a *move-prf* template to provide movement of a proform along a path. Other proform templates will be added as the system matures. Figures 2 and 3 show two frames superposed from prform movements. The movement in in Figure 2 corresponds to the AZee expression

Figure 2: A standing person moving to center

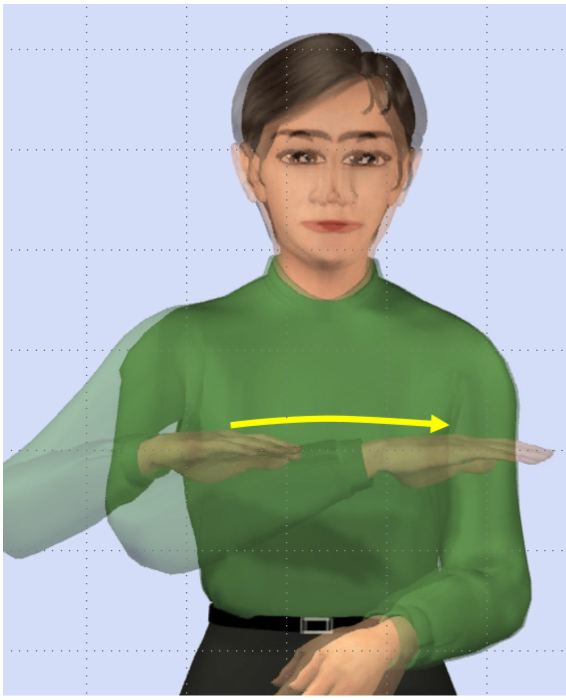


Figure 3: A vehicle moving from left to right

move-prf(proform-standing-person,
path(straight, rssp, midssp))

whereas the movement in Figure 3 corresponds to the LSF expression

move-prf(proform-vehicle, path(straight, rssp, lssp))

(a) place-prf(proform-standing-person, rssp)

(b) place-prf(proform-vehicle, midssp)

Figure 4: A standing person on the right of the signing space (a); a vehicle in the center (b)

The slight blur in the torso and face in 2 is indicative of the subtle spine movement that automatically supports the motion. Larger arm motions across the body as in Figure 3 are naturally accompanied by a larger torso motion, just as they would be for a live signer. This avatar torso movement is an improvement over the results of prior efforts in which the torso was stationary, and contributes to greater naturalness in the resulting animation.

Figures 4 and 5 demonstrate the generated placement movement for a selection of proforms and placement sites. The system can use any site in front of the body for placement and movement, both for one-handed and two-handed proforms. These figures further illustrate the responsiveness in the torso algorithm through the degree of blur in the head caused by the superposition of the two frames. The torso reacts more to the two-handed motion in figures 5a and 5b, causing a larger blur relative to the motion shown in figures 4a and 4b.

An animated version that demonstrates placement and movement of a selection of proforms is available online at `http://asl.cs.depaul.edu/proforms/proformPlacementAndMovement.mp4`. This animation also provides examples of placement in several locations to show the flexibility of the system. In all of these cases, Paula uses its procedural models of human motion along with data from the artist template to provide natural poses and transitions.

4. Conclusion and future work

This paper represents a first attempt at achieving more natural proform placement and movement using the structured linguistic model AZee to drive the hybrid animation system Paula. By extending the existing XML matching scheme for shortcutting to allow templating on the functional AZee expression, the system provides a flexible way for the animation system to leverage animator specified data for known proforms to improve postures and movement, while specifying a robust fallback algorithm, that still requires implementation. This sacrifices quality when the proform actions or the proforms themselves are not recognized.

Moving forward, the bridge currently supports a selection of movement procedures, but more work needs to be done both linguistically and geometrically to identify other proform structures that can be shortcut. In particular, the present work has focused on placement and straight movement of isolated one and two-handed proforms. More generally, proforms in sign languages exhibit a wide range of motion styles including bounce and wavy motions that indicate styles of movement, very general spatial movements to trace the size and shapes of objects, and relative proform placement and movement for complex scene descriptions. The current models have been built to support many of these structures, but extended study is needed for such complex motions to both refine the linguistic descriptions and naturally coordinate motion between interacting proforms.

(a) place-prf(proform-flat-round-small, rssp)

(b) place-prf(proform-flat-round-small, lssp)

Figure 5: A plate placed on the right of the signing space (a); a plate placed on the left (b)

Finally, in the future, we hope to set up a web interface providing rendered videos from input AZee expressions. This way, users will be able to connect and sandbox with our system.

Acknowledgements

The authors would like to thank Université Paris Sud for their support of Dr. John McDonald through their "recrutement de professeur invité" grant program, Dr. Annelies Braffort for her valuable input and suggestions concerning proforms and Dr. Rosalee Wolfe for her valuable input on avatar posture.

5. Bibliographical References

Braffort, A., Benchiheub, M., and Berret, B. (2015). Aplus: A 3d corpus of french sign language. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 381–382. ACM.

Elliott, R., Glauert, J. R., Kennaway, J., Marshall, I., and Safar, E. (2008). Linguistic modelling and language-processing technologies for avatar-based sign language presentation. *Universal Access in the Information Society*, 6(4):375–391.

Filhol, M. and Braffort, A. (2006a). A sequential approach to lexical sign description. In *Proceedings of LREC 2006, second workshop on the representation and processing of sign languages: lexicographic matters and didactic scenarios*, Genova, Italy.

Filhol, M. and Braffort, A. (2006b). Sign description: how geometry and graphing serve linguistic issues. *Theoretical issues on Sign Language research, TISLR*, 9.

Filhol, M., Hadjadj, M. N., and Choisier, A. (2014). Non-manual features: the right to indifference. In *Representation and Processing of Sign Languages: Beyond the manual channel, Language resource and evaluation conference (LREC)*, Reykjavik, Iceland.

Filhol, M., McDonald, J., and Wolfe, R. (2017). Synthesizing sign language by connecting linguistically structured descriptions to a multi-track animation system. In *International Conference on Universal Access in Human-Computer Interaction*, pages 27–40. Springer.

Gibet, S., Courty, N., Duarte, K., and Naour, T. L. (2011). The signcom system for data-driven animation of interactive virtual signers: methodology and evaluation. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 1(1):6.

Huenerfauth, M. (2004). Spatial representation of classifier predicates for machine translation into american sign language. In *Workshop on Representation and Processing of Sign Language, 4th Internationnal Conference on Language Ressources and Evaluation (LREC 2004)*, pages 24–31.

Liddell, S. K. (2003). *Grammar, gesture, and meaning in American Sign Language*. Cambridge University Press.

Lombardo, V., Battaglino, C., Damiano, R., and Nunnari, F. (2011). An avatar-based interface for the italian sign language. In *Complex, Intelligent and Software Intensive Systems (CISIS), 2011 International Conference on*, pages 589–594. IEEE.

López-Colino, F. and Colás, J. (2011). The synthesis of lse classifiers: From representation to evaluation. *Journal of Universal Computer Science*.

McDonald, J., Wolfe, R., Alkoby, K., Carter, R., Davidson, M. J., Furst, J., Hinkle, D., Knoll, B., Lancaster, G., Smallwood, L., et al.). Achieving consistency in an fk/ik interface for a seven degree of freedom kinematic chain. In *Visual Communication and Technology Education Faculty Publications. 3*, pages 171–179.

McDonald, J., Wolfe, R., Schnepp, J., Hochgesang, J., Jamrozik, D. G., Stumbo, M., Berke, L., Bialek, M., and Thomas, F. (2016a). An automated technique for real-time production of lifelike animations of american sign language. *Universal Access in the Information Society*, 15(4):551–566.

McDonald, J., Wolfe, R., Wilbur, R. B., Moncrief, R., Malaia, E., Fujimoto, S., Baowidan, S., and Stec, J. (2016b). A new tool to facilitate prosodic analysis of motion capture data and a data-driven technique for the improvement of avatar motion. In *Proceedings of Language Resources and Evaluation Conference (LREC)*, pages 153–59.

McDonald, J., Wolfe, R., Johnson, S., Baowidan, S., Moncrief, R., and Guo, N. (2017). An improved framework for layering linguistic processes in sign language generation: Why there should never be a brows tier. In *International Conference on Universal Access in Human-Computer Interaction*, pages 41–54. Springer.

Weast, T. (2011). American sign language tone and intonation: A phonetic analysis of eyebrow properties. *Formational units in sign languages*, pages 203–228.

Wolfe, R., McDonald, J., and Schnepp, J. (2011). Avatar to depict sign language: Building from reusable hand animation. *International Workshop on Sign Language Translation and Avatar Technology (SLTAT)*.

Building French Sign Language Motion Capture Corpora for Signing Avatars

Sylvie Gibet
Université Bretagne-Sud, Laboratoire IRISA, France
sylvie.gibet@univ-ubs.fr

Abstract

This paper describes four corpora that have been designed and built in our research team. These corpora have been recorded using motion capture (MoCap) and video equipment, and annotated according to multi-tiers linguistic templates. Each corpus has been designed for a specific linguistic purpose and is dedicated to data-driven synthesis, by (i) replacing signs or groups of signs within an utterance, (ii) replacing phonetic or phonological components and in this way modifying the grammatical or semantic aspects of the phrase, or (iii) altering prosody in the produced sign language utterances.

Keywords: Corpus, MoCap, Sign Language, Signing avatar

1. Introduction

The design of traditional corpora for linguistic analysis aims to provide living representations of sign languages across deaf communities and linguistic researchers. Most of the time, the sign language data is video-recorded and then encoded in a standardized and homogenous structure for open-ended analysis (statistical or phonological studies). With such structures, sign language corpora are described and annotated into linguistic components, including phonology, morphology, and syntactic components (Johnston and de Beuzeville, 2009; Crasborn and Zwitserlood, 2008; Efthimiou and Fotinea, 2007; Wolfe et al., 2011; Hanke et al., 2012).

Conversely, motion capture (MoCap) corpora provide researchers the data necessary to carry on finer-grained studies on movement, thus allowing precise, and quantitative analysis of sign language gestures as well as sign language (SL) generation. One the one hand, motion data serves to validate and enforce existing theories on the phonologies of sign languages. By aligning temporally motion trajectories and labelled linguistic information, it thus becomes possible to study the influence of the movement articulation on the linguistic aspects of the SL, including hand configuration, hand movement, co-articulation or synchronization within intra and inter phonological channels. On the other hand, generation pertains to sign production using animated virtual characters, usually called signing avatars.

Although MoCap technology presents exciting future directions for SL studies, tightly interlinking language components and signals, it still requires high technical skills for recording, post-processing data, and there are many unresolved challenges, with the need to simultaneously record body, hand motion, facial expressions, and gaze direction.
Therefore, there are still few MoCap corpora that have been developed in the field of sign language studies. Some of them are dedicated to the analysis of articulation and prosody aspects of sign languages, whereas recent interest in avatar technology has led to develop corpora associated to data-driven synthesis. In particular, (Lu and Huenerfauth, 2014) collected an ASL corpus and discussed how linguistic challenges in ASL generation could be addressed through this corpus. To improve avatar movement, kinematic and linguistic cues were retrieved from motion capture data and incorporated into a data-driven technique, thus leading to a more realistic animation (Mcdonald et al., 2016). More recently, a MoCap dataset on French sign language (LSF) has been collected (Limsi and CIAMS, 2017). 25 pictures are described in a spontaneous way, and first analysis are conducted. However these corpora do not capture simultaneously the multiple channels involved in SL gestures, i.e. body movement, hand configuration, facial expression, and gaze direction, which remains an important challenge and is necessary to address these highly coded sign languages using multi-tiers linguistic elements, both for recognition (Dilsizian et al., 2014) or synthesis (Gibet et al., 2011).

In this article, we describe four motion capture corpora in French sign language that were designed and built in our research team during the last decade. The technical aspects of the MoCap acquisition are described. Then the linguistic rules that guide the corpora design for the synthesis of new sentences in LSF are discussed and illustrated with examples. Most of these linguistic issues are applied to data-driven generation. However, in all our studies, it is important to emphasize that we adopted a synthesis-by-analysis approach. That is to say, the improvement of our synthesis models led us progressively to refine our methods of segmentation and labeling, to better understand the mechanisms responsible for the formation of the signs, as well as the processes of coarticulation (Naert et al., 2017).

2. Motion Capture Datases

Four corpora in French Sign Language (LSF) and their corresponding MoCap databases have been designed by a team of researchers that includes linguists and computer scientists, Hearing and Deaf. Before describing these corpora we describe hereafter the motion capture databases that were collected over the last ten years in the context of national research projects, with different objectives concerning the linguistic aims and the level of avatar synthesis.

Projects	Markers Capture device	Cameras number	frequency Hz	Databases	Year	Size min
HuGEx	24 (body) 2 Cybergloves 39 (face)	12	120	TRAIN METEO	2005	10' 40'
SignCom	43 (body) 2x6 (hands) 41 (face)	12	100	SignCom	2009	60'
Sign3D	40 (body) 2x19 (hands) 40 (face) gaze direction	16	100	Sign3D	2013	10'

Table 1: MoCap databases built at IRISA.

2.1. Mocap Devices and Experimental protocols

Different MoCap setups and experimental protocoles were defined in the context of three projects. For all of them, we used a Vicon MX infrared camera technology to capture the 3D displacements of a set of markers. The main differences between the setups of the projects are the number of cameras, of markers, and the frequency of acquisition. Table 1. gives an overview of the projects with the experimental setups. For capturing precisely movements of the hands and facial expressions, it is necessary to use more cameras and to place them closer to the subject so that they can detect all the markers. This is why we increased the number of cameras as we gained experience and mastered the motion capture systems. In addition, the frequency of acquisition has to be large enough to capture the subtle variations of the movements, for example when changing a facial expression, or moving rapidly one hand. We therefore tried to determine the appropriate frequency of acquisition, trying to keep a good compromise between the spatial accuracy and the speed of the cameras. Finally, in all our experimental setups, we considered pairing MoCap with video recordings, assuming that parallel recordings would aid in the ulterior data annotation processes.

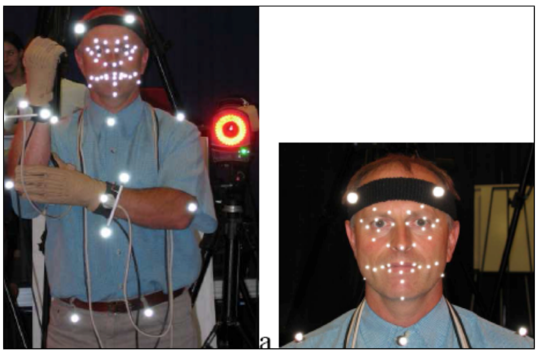


Figure 1: Photo of the MoCap settings in *HuGEx* project.

2.1.1. *HuGEx* project

In *HuGEx* project (Gibet et al., 2006), the Vicon system was composed of 12 infrared cameras cadenced at 120Hz. For the body movements 24 reflective markers were placed on standardized anatomical landmarks. We also recorded facial expressions using 39 small semi-spherical markers (3mm) at locations compliant with Mpeg4 specification. As we had no experience with hand capture, hand movements were recorded using two Cyber gloves (Ascension technologies), each one composed of 22 sensors (see Fig. 1). The fusion of the different signals (body, left and right hand) was realized after reconstruction and synchronization (resampling at 60 Hz). The different information sources (body and hands) were then converted into BVH format. During the recording session, about forty minutes of LSF gestures were captured on one expert deaf signer. This one, who was a trainer in LSF, signed on texts that he himself transcribed into a sequence of glosses. Two databases were built: (i) the *TRAIN* database (about 10 min), aimed at building sentences with predefined replaceable parts; (ii) the *METEO* database (about 40 min), aimed at studying the variation in prosody of the LSF phrases. For both datasets, the mean duration of a sequence was about 60 seconds.

Figure 2: Photo of the MoCap settings in *SignCom* project.

2.1.2. *SignCom* project

The *SignCom* project (Gibet et al., 2011) also used a Vicon MX system with 12 high definition cameras to capture the movements of our LSF signers at a frequency rate of 100Hz. We had 43 markers for the body, and 41 markers of small diameter for the face. Instead of using data gloves which lack precision and exhibit significant drift when used for a long time, we captured the hand movements by fixing 6 markers per hand (see Figure 2).
As for the previous projects, we used an additional video camera to have video recordings in addition to MoCap

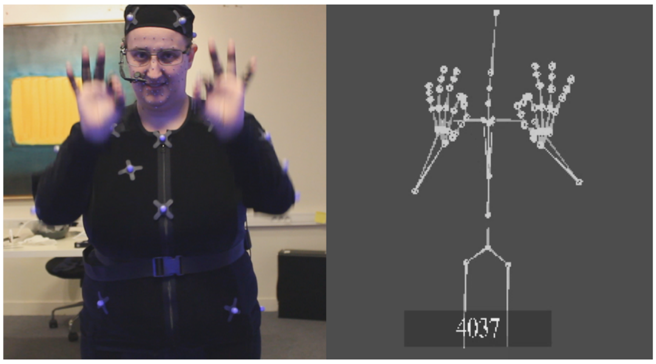


Figure 3: Photo of the MoCap settings in *Sign3D* project.

data. This is necessary for annotation. Body movements, hand and finger movements, and facial expressions were recorded simultaneously by the Vicon system. Two professional deaf linguists signing in LSF designed the corpus, and learned it by heart. During the recording session, the information was presented to the deaf signers through images projections, so that the signers were able to recall the scenarii without reading any text-based translation. 68 motion sequences were recorded on one signer. This constitutes the *SignCom* database containing about one hour of MoCap data. From this data (in C3D format), a skeleton was reconstructed, and the data (body and hand movements, as well as facial expressions) was converted into the formats BVH and FBX.

2.1.3. *Sign3D* project

The *Sign3D* project (Gibet et al., 2015) used a Vicon T160 system with 16 high definition cameras at a frequency rate of 100 Hz, combined with a head-mounted oculometer (MocapLab MLab 50-W), designed to track the gaze direction. Facial expressions, body and finger motions were again simultaneously recorded. For recording precisely hand movements and hand configurations, we used a much larger number of markers (19 per hand against 6 in the *SignCom* project). The LSF gestures of one expert deaf signer were recorded for about 10 minutes to form the dabase *Sign3D*. The motion capture settings associated to the skeleton reconstruction is illustrated in Figure 3. The motion skeleton data including body and hand movements) was converted into the FBX format. The facial expressions data was converted into blendshape coefficients.

3. Corpora Design

We describe hereafter the various corpora that we have designed in the context of the former research projects. Through these projects, we developed a complete concatenative data-driven synthesis pipeline that enables the assembling of motion elements, from signs and parts of sentences, to motion chunks retrieved from different channels and body parts (hand movements, hand configurations, body movements, facial expressions, and gaze direction), representing phonetic or phonological components.

Our corpora follow the objectives of synthesis by replacing signs or groups of signs in sentences, by composing phonetic or phonological components, and finally by analyzing and generating prosody in sentences carried out in different stylistic contexts.

3.1. Motivation

For the purpose of corpus design, three main questions have been addressed in the three former projects. The first one concerns the corpus content itself and the compromise that exists between breadth and depth in its design. The second question concerns the nature of the sign variability which is of paramount importance if we want to create new sentences in different discourse contexts. The third question concerns the acted or spontaneous nature of the produced SL utterances.

Concerning the first question, we wanted to have control over the signs or gloses that appeared into the corpora, and therefore we would prefer a limited vocabulary of given signs, and multiple instances for each sign played in different discourse contexts. We also chose to incorporate standard signs into our lexicon, as they were easier to handle for synthesis. Given the difficulty of capturing large corpora (a tedious and time-consuming process, both in terms of capture, post-processing, and annotations), we also opted for a limited set of utterances or sequences of signs. Therefore, in parallel with the design of our sentences, we had to think deeply about the mechanisms of editing signs and constructing new sentences.

The question of variability can be approached in different ways: (i) by constructing sentences containing the same signs appearing in different contexts; (ii) by repeating the sentences several times and with different subjects; and (iii) by enriching the initial corpus with new constructed sentences.

To answer the third question, in all our corpora, the scenarii were scripted by deaf persons, and the produced sign language utterances were acted. Table 2 gives an overview of the corpora, indicating the level of annotation, the topic, and the linguistic application.

3.2. Replacing Signs or Groups of Signs in Sentences

The first experimental ideas for synthesizing new statements from original sign language data were to insert replaceable parts into a sentence, such as signs, or groups of signs. This was first achieved in the *HuGEx* project where the corpus was composed of a set of phrases expressing incident reports relatively to the railway traffic, with a set of additional signs representing French towns. Two excerpts are shown below. The brackets delimit the variable elements.

> The train from [CITY] to [CITY] is delayed by [DURATION], due to [CAUSE].
> The train [NUMBER] is being prepared; the starting lane will be displayed in [DURATION].

It was then possible to build programmable sentences by choosing the departure and arrival cities ([CITY]) among a given set of pre-recorded cities, the number of the train ([NUMBER]), or the nature of the incident ([CAUSE]) belonging to the following set: a technical incident / bad

Databases	Annotation, Segmentation	Nature	Linguistic Application
TRAIN	gloss	Train incident Towns, numbers	Fill-gap-synthesis
Sign3D	phonetic, phonology, gloss	Urban services Places, schedules, event rates	Phonological synthesis Hand movement analysis
SignCom	phonetic, phonology, gloss	Recipes Interactive dialogs	Pattern-based synthesis Coarticulation analysis
METEO	gloss	Weather forecast Emotional variations	Prosody analysis

Table 2: Corpora in LSF.

weather / personal accident).

In the *Sign3D* project, we collected a corpus of French sign language utterances, describing various events (exhibitions, inaugurations, cultural events) taking place in various buildings and monuments (swimming pool, theater, town hall, museum, etc.), indicating the opening and closing hours, entrance fees, their location relative to each other, and the potential occurrence of an incident (weather, work, etc.). In this latter corpus, the aim was also to build new sentences by replacing signs (hours, buildings, etc.), or groups of signs (events, incident causes). To preserve the linguistic coherence of the LSF statements, while optimizing the number of variants of the different sentences, the corpus was designed by declining a limited set of syntactic patterns (the brackets delimit the variable elements).

> The [LOCATION] is [ABSOLUTE OR RELATIVE LOCATION]; it opens at [TIME] and closes at [TIME].
> Access is [PAYING / FREE], [ENTRY CONDITION].
> The [EVENT] in [LOCATION] is moved to [LOCATION] due to [CAUSE].
> In case of [CAUSE], the [EVENT] in [LOCATION] will be moved to [LOCATION].

where the variables ([LOCATION], [TIME], etc.) may be replaced by values belonging to a given set of signs. Thirteen sample sentences were then signed by a deaf LSF expert. This corpus can easily be extended by enriching it with the synthesized variant sentences. This represents about 10 minutes of continuous LSF.

3.3. Altering Phonological Components of Signs

The objective of the *SignCom* project was also to design new utterances in LSF, and to animate a virtual signer, using both raw motion and annotated data. The idea was similar to the previous projects, but instead of manipulating signs, the aim was to re-assemble phonetic or phonological elements of signs, while keeping the global coherence and realism of the produced sequences. The corpus contains three thematic scenarii: the *Cocktail* monologue, and the *Galette* and *Salad* dialogues. The scenarii were scripted by two expert deaf people who designed the scenes using comic stories that were displayed on the back wall of the room, thus avoiding lowering the head for reading the scenarii. Both deaf people trained for several days before the recording sessions, hence they executed the motion as acted sequences. A total of 68 sequences was captured and annotated, following a multi-tier template with different levels of labeling (gloss, phonological and phonetic elements), for each hand separately, and for the two-hands.
As signed languages are by nature spatial languages, forming sign sequences requires a signer to understand a set of spatial-temporal grammatical rules and inflection processes. These processes have oriented the range of LSF signs recorded for the project. This brought us to include a number of various linguistic inflection mechanisms into the corpus that allow for creating novel sentences from our original corpus. After defining a delimited vocabulary, we chose to introduce spatial references (for example depicting and indicating verbs which are modulated in the context of dialog situations), changes in hand configurations, and changes in hand movements.

3.3.1. Spatial references: directional verbs and pointing movements

We included in the dataset directional verbs and depicting verbs as well as personal and possessive pronouns. This gave us the possibility to build new sentences by conjugating the verbs. For example, the sign INVITE can be modified grammatically to become ”I invite you”, ”You invite me”, etc. Our vocabulary thus contains several instances of the directional verbs shown in Table 3. A certain number of pointing gestures in different parts of the signing space were also included. These targets are labeled with their 3D location.

3.3.2. Changes in hand configurations

Many hand configurations, possibly associated to verbs, allow for designing different objects, or indicate *size* or

Salad	Cocktail	Directional verbs
22 × SALAD	8 × COCKTAIL	GIVE
20 × PRO-1	8 × DRINK	TAKE
19 × WHAT	7 × GLASS	PROPOSE
8 × PLATE	7 × FRUIT	INVITE
6 × TOMATO	3 × ORANGE	COMMUNICATE
12 × POUR	3 × JUICE	PUT
11 × WANT	7 × FILL	EXPLAIN
9 × CHEVRE	7 × THERE-IS	QUESTION
9 × VARIOUS	3 × NEXT	
3 × AVOCADO	4 × ALCOHOL	
5 × ADD	2 × WITHOUT	

Table 3: Two first columns: some tokens with their occurrence in the *Salad* and *Cocktail* scenarii (*SignCom* corpus); Third column: directional verbs mainly found in the dataset.

shape specifiers. Given our inclusion of signs that take multiple hanshapes, like GIVE, we introduced in the corpus different hand configurations from other signs that can be substituted to the original handshapes. In the case of GIVE, most often signed in our dataset as if the signer was handing a glass to someone, a hanshape substitution could yield addition meanings, such as giving a piece of paper or giving an object with a cylindrical shape. In particular, the expression GIVE A GLASS is performed in our corpus with glasses of different sizes and forms (for example a large glass, a thin long glass, or a champagne flute).

3.3.3. Changes in movement kinematics

Analyzing hand movements has shown regular shapes (bell-shapes) which differ whether they belong to strokes (within-sign) or transitions (inter-signs). Moreover, for strokes, toward-target movements differ from backward movements (Duarte and Gibet, 2010). These observations have led us to introduce many kinematic variations of movements in the corpus, so that it becomes possible to analyze and annotate these patterns, and to retrieve the appropriate movement from the database that preserve the temporal coherency of the reconstructed phrase.
The corpus also contains reversal verbs, as for example the sign GIVE which can be reversed in the sign TAKE, or the sign LIKE which can be reversed in DO-NOT-LIKE in LSF.

3.3.4. Composition process to build new sentences

An overview of the most frequent tokens in the Cocktail and Salad scenarii is provided in Table 3. With this variety and frequency of our related lexemes, we are able to produce a number of novel utterances based on the thematic subjects. Examples of construction of new sentences from the above transformations is shown in the following examples:

> I GIVE-YOU a THIN-GLASS (1)
> I TAKE a LARGE-GLASS (2)
>
> I LIKE FRUIT JUICE (3)
> I DO-NOT-LIKE ORANGE JUICE (4)

In this first example, only the right arm is involved. The movement (2) begins at the position where the movement (1) ends; the direction of movement is reversed, and the shape of the hand is changed to handle a big glass instead of a thin glass. In the second example, different channels are combined, by keeping the torso/lower-body/left-arm of one sequence (3), and substituting the head, facial expression and right arm movements of another sequence (4). The sign DO-NOT-LIKE is reversed from the sign LIKE. In such a composition process, the spatial constraints should be preserved, in particular the sign ORANGE should be executed near the corresponding body part (head), whatever the torso or the head orientation is. This clearly reveals that the combination process should be driven at a more abstract level, expressed by rules or constraints incorporated into the animation engine.

3.4. Altering the Prosody of Sentences

The objective of the *HuGEx* project was to animate with a data-driven approach a virtual signer endowed with expressive sign language gestures. Our attention focused on the prosody of the LSF gestures, and on its influence on the semantic comprehension. The corpus *METEO* was composed of a set of sentences describing weather forecasts, performed with different variations of expressiveness: *neutral*, *angry*, *emphasis*, and *tired*. The mean duration of a sequence was 60 seconds. We took as referent performance the first sequence performed according to neutral style. An example retrieved from the corpus is given below.

> Today, July 6th, here is the weather forecast. In the morning, clouds will cross Brittany. In the afternoon, it will rain. Tomorrow, the sun will shine. It will be hot and dry.

An analysis of the LSF prosody was achieved on the expressive sentences, through a temporal alignment process using an adaptive dynamic time warping algorithm (Héloir et al., 2006). Using machine learning techniques, it would be possible to learn the sequences performed with different styles and then transfer the style of one sequence into another one.

4. Conclusion

In this article, we described four corpora with different linguistic purposes that have been designed and built in our research team over the last ten years. These corpora were recorded using MoCap data, post-processed and manually annotated, so that they could be used for different goals: linguistic analysis, automatic annotation, or generation.

With the increasing interest of linguistic or computer science researchers using sign language motion capture data, there is a need to provide motion capture databases that can be shared by the different communities. Following the approach adopted by other research teams in movement sciences that have made available raw motion, videos, and tools, with exchangeable data formats, we want to share our experience on the design of corpora and the construction of MoCap databases. We also propose to make available soon our LSF MoCap corpora.

Concerning the annotated data, we have used schemes inspired from the linguistic community, and we are currently enriching these schemes by developing automatic annotation methods. These annotated schemes (manual or automatic) with the documentation explaining the structure and the coding system of the annotation should also be shared by the different research communities.

Other more focused corpora are also currently designed and collected in our research team. They are dedicated to the automatic annotation of two specific channels: facial expressions and hand configurations, and will be used for animating a signing avatar (Naert et al., 2018).

5. Acknowledgements

The motion capture datasets presented in this paper have been realized, thanks to the *HuGEx* project funded by the CNRS, the *SignCom* project funded by the French National Research Agency (ANR), and the *Sign3D* project, funded by the French Ministry of Industry.

6. References

Crasborn, O. and Zwitserlood, I. (2008). Annotation of the video data in the Corpus NGT. Technical report, Department of Linguistics and Center for Language Studies, Radboud University, Nijmegen, the Netherlands, November.

Dilsizian, M., Yanovich, P., Wang, S., Neidle, C., and Metaxas, D. (2014). A new framework for sign language recognition based on 3d handshape identification and linguistic modeling. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

Duarte, K. and Gibet, S. (2010). Reading between the signs: How are transitions built in signed languages? In *Theretical Issues in Sign Language Research (TILSR 2010), Indiana, USA*.

Efthimiou, E. and Fotinea, S.-E. (2007). GSLC: Creation and annotation of a Greek Sign Language corpus for HCI. In *Universal Access in Human Computer Interaction. Coping with Diversity*, volume 4554 of *Lecture Notes in Computer Science*, pages 657–666. Springer, Berlin.

Gibet, S., Héloir, A., Courty, N., Kamp, J., Gorce, P., Rezzoug, N., Multon, F., and Pelachaud, C. (2006). Virtual agent for deaf signing gestures. In *AMSE, Journal of the Association for the Advancement of Modelling and Simulation Techniques in Enterprises (Special edition HANDICAP)*, pages 127–136.

Gibet, S., Courty, N., Duarte, K., and Le Naour, T. (2011). The signcom system for data-driven animation of interactive virtual signers : Methodology and evaluation. In *Transactions on Interactive Intelligent Systems*, volume 1. ACM.

Gibet, S., Lefebvre-Albaret, F., Hamon, L., Brun, R., and Turki, A. (2015). Interactive editing in French sign language dedicated to virtual signers: requirements and challenges. *Universal Access in the Information Society*, 15(4):525–539.

Hanke, T., Matthes, S., Regen, A., and Worseck, S. (2012). Where does a sign start and end? Segmentation of continuous signing. *Language Resources and Evaluation Conference*.

Héloir, A., Courty, N., Gibet, S., and Multon, F. (2006). Temporal alignment of communicative gesture sequences. *Computer Animation and Virtual Worlds*, 17:347–357.

Johnston, T. and de Beuzeville, L. (2009). Researching the linguistic use of space in Auslan: Guidelines for annotators using the Auslan corpus. Technical report, Department of Linguistics, Macquarie University, Sydney, June.

Limsi and CIAMS. (2017). Mocap1. ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr.

Lu, P. and Huenerfauth, M. (2014). Collecting and evaluating the cuny asl corpus for research on american sign language animation. *Comput. Speech Lang.*, 28(3):812–831, May.

Mcdonald, J., Wolfe, R., Wilbur, R., Moncrief, R., Malaia, E., Fujimoto, S., Baowidan, S., and Stec, J. (2016). A new tool to facilitate prosodic analysis of motion capture data and a data- driven technique for the improvement of avatar motion. In *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining; Language Resources and Evaluation Conference*, volume 7.

Naert, L., Larboulette, C., and Gibet, S. (2017). Coarticulation analysis for sign language synthesis. In *International Conference on Universal Access in Human-Computer Interaction*, pages 55–75. Springer.

Naert, L., Reverdy, C., Larboulette, C., and Gibet, S. (2018). Per channel automatic annotation of sign language motion capture data. In *8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community; Language Resources and Evaluation Conference (LREC)*, Miyazaki, Japon, May 12.

Wolfe, R., McDonald, J., Schnepp, J., and Toro, J. (2011). Synthetic and acquired corpora: Meeting at the annotation. In *Workshop on Building Sign Language Corpora in North America, Washington, DC*.

Modeling of geographical location in French sign language from a semantically compositional grammar

Mohamed Nassime Hadjadj
LIMSI, CNRS, Universite Paris Saclay
Rue John Von Neumann, 91400 Orsay, France
mohamed.hadjadj@limsi.fr

Abstract

The use of the specificities related to the visuo-gesual modality of SL, such as the use of the signing space and the simultaneous articulation of multiple channels allows the signer to express structures in a more illustrative way. The description of this structure goes beyond the linear linguistic organization initially applied to describe spoken languages. In this paper, we are interested in modeling structures that rely on the signing space to designate the location of one object relative to another. We are particularly interested in the study of location of one place in relation to another one in French Sign Language (LSF). After a presentation of the corpus and the methodology followed to analyze it, we present the study carried out as well as the results obtained.

Keywords: French sign language, computer modelling, formal grammar

1. Introduction

Many language studies on sign language (SL) have observed that some of the structures are not defined as a linear sequence of signed units (Woll, 2007), (Filhol et al., 2014). The specificities related to the visuo-gesual modality of SL, such as the use of the signing space and the simultaneous articulation of multiple channels, allow the signer to express structures in a more illustrative way. The description of this structure goes beyond the traditional linguistic organization initially applied to describe the spoken languages (Huenerfauth, 2006), (Cuxac and Sallandre, 2007). In this paper, we rely on an empirical approach to model structures that use signing space to designate the location of one object relative to another. We are particularly interested in the location of one place in relation to another in French Sign Language (LSF).We briefly present in this paper the methodology we used to model some structures from a corpus study. We then present the studied corpus and the results obtained.

2. Methodology

To model the localization structures present in corpora, we adopt an approach that moves away from the constraints imposed by linear models (Hadjadj et al., 2018). In other words, we do not suggest any linguistic organization of SLs beforehand. Our approach aims to identify, from LSF corpora, a set of “observable forms” that refers to the same “semantic function”. The notion of observable form defines gestural articulations and the different synchronizations that can take place between them. If we take as an example the following articulations:

A: “an eye blink”

B: “move the index finger down”

Each of these two gestural articulations can be defined as a form. Also, a form can be composed simultaneously of several articulations (A and B for example). In this case, the different synchronizations that characterize them are considered as a form criterion. By the notion of semantic function, we mean the interpretation attributed to an observable form. The following examples are considered as possible semantic functions:

C: the concept of “House”

D: Negative expression on a variable element

E: location of an object (obj1) in relation to another object (obj2)

Any systematic association between a single group of observable forms and a semantic function is considered as a rule that participates in the linguistic organization of the language. This article is part of a project of LSF generation. Thus, we are interested in identifying production rules (function-to-form links). It should also be noted that the identification of a link between a single form group and a semantic function may require several function-to-form iterations. We define a production rule by the following triple:

- An identifier: usually the name of the semantic function of the rule
- Arguments of the rule: the set of parameters on which the rule may depend
- Associated form: the invariant forms of the rule and their possible dependencies on the arguments.

Box diagrams (figure 1) can illustrate a production rule where:

- The horizontal axis represents the production time.
- The boxes represent time intervals in which an articulation must take place. The articulators are set in bold; their positions are designated in italics
- The blue boxes are invariant specifications.
- The boxes in red represent the time intervals during which an argument is to be produced.

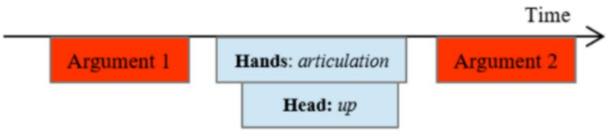


Figure 1: Example of a production rule

3. Study and results

In order to study LSF structures that do not necessarily respect a linear order, we start this study with a criterion L: the geographical location of one place in relation to another. The expression of a link between two places or objects may require a more complex representation than a linear sequence of signed units (Lejeune, 2004). We present below the corpus used and the study carried out

3.1 Corpus

The corpus "websourd AFP 2007" consists of 2000 short summaries of AFP newswire articles of the year 2007. It was signed throughout 2007 by the signers of the company Websourd and it covers various topics: economy, politics, health etc. The large number of newswire articles ensures a relevant number of occurrences of the same linguistic phenomenon as well as its production by several signers.

3.2 First iteration L

Starting this study with a first iteration of function form, we identified three groups of forms. Thus, in our approach, each identified group of occurrences becomes a starting criterion for a new iteration. We stop this process once we define a link between a semantic function and a single form group. We present below the different steps of this study

Iteration L: Location of a place 1 in relation to a place 2 (function criterion)

Example: "Tens of thousands of Shiites arrived on Monday in Najaf (160 km south of Baghdad)" (place 1: Najaf, place 2: Baghdad)

- Number of occurrences in the corpus (Nocc) = 147
- Number of groups identified (of form in this iteration) (Ngp) = 3
- Occurrences that do not fit into any group (Nout) = 15

Group L.1:

Figure 2: pointing sign Figure 3: Articulation of both hands

Figure 4: Location

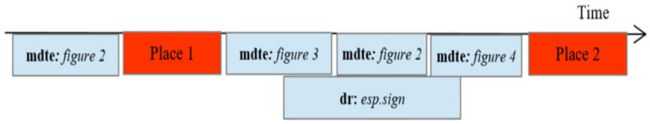


Figure 5: Form of group L.1

The form of group L.1 is composed of:

- Articulation of the strong hand: figure 2
- Argument: place 1
- Articulation of the weak (mde) and strong hands - (mdte): figure 3
- Articulation of the strong hand (mdte): figure 2
- Articulation of the weak (mde) and strong hands (mdte): figure 4
- Eye gaze directed to the signing space (dr: esp-sign)
- Argument 2: place 2

Group L.2:

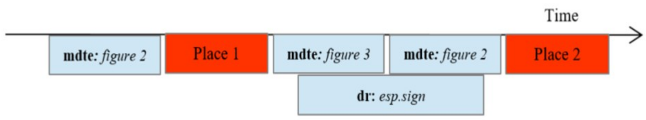


Figure 6: Form of group L.2

The form of group L.2 is composed of:

- Articulation of the strong hand: figure 2
- Argument: place 1
- Articulation of the weak (mde) and strong hands (mdte): figure 3
- Articulation of the strong hand (mdte): figure 2
- Eye gaze directed to the signing space (dr: esp-sign)
- Argument 2: place 2

Group L.3

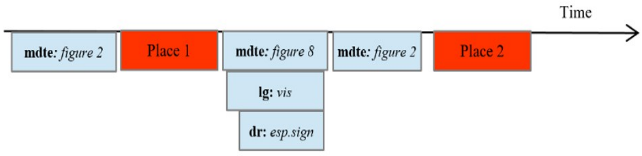


Figure 7: Form of group L.3

Figure 8: Near

The form of group L.3 is composed of:

- Articulation of the strong hand: figure 2
- Argument: place 1
- Articulation of the weak (mde) and strong hands (mdte): figure 8
- The tongue of signer: vsible (lg : vis)
- Eye gaze directed to the signing space (dr: esp-sign)
- Articulation of the strong hand: figure 2
- Argument 2: place 2

3.3 New iterations from L.x groups

We present in the following sections the different iterations made from the three form groups identified during the first iteration as well as the defined production rules.

Form criterion L.1: cf.fig.5

- Nocc = 70
- Ngp = 1
- Nout = 5

Single group: place 2 is a part with undefined borders within place1

Example: place 1 "France", place2 "south of France"

Function criterion L.1.1. Place 2 is a part with undefined borders within place1

- Nocc = 65
- Ngp = 1
- Nout = 0

The condition of our methodology is verified, the iteration starts with a function criterion associated with a unique group of forms. This defines a production rule, specified as follows:

Production rule L1.1:

- **Identifier**: Place 2 is a part with undefined borders within place1
- **Arguments**: place1, place 2
- **Form**: see Figure 5

Form criterion L2: cf.fig.6

- Nocc = 33
- Ngp = 1
- Nout = 6

Single group: place 2 is a part with defined boundaries inside of place 1

Example: place 1 " France ", place 2 " Marseille"

Function criterion L.2.1: Place 2 is a part with defined boundaries of inside place 1

- Nocc = 27
- Ngp = 1
- Nout = 0

Production rule L2.1:

- **Identifier**: Place 2 is a part with defined boundaries inside of place 1
- **Arguments**: place1, place2
- **Form**: see Figure 6

Form criterion L.3: cf.fig.7

- Nocc = 29
- Ngp = 1
- Nout = 4

Single group: place 1 is near place 2

Example: "Heads of G8 diplomacy meet on Wednesday in Potsdam near Berlin to prepare the international agenda for the Heiligendamm Summit (6-8 June)". Place 1: Potsdam is near place 2: Berlin.

Function criterion L3.1: place 2 is near place1

- Nocc = 25
- Ngp = 1
- Nout = 0

Production rule L3.1:

- **Identifier**: Place 2 is near place1

- **Arguments**: place1, place2
- **Form**: see Figure 7

3.4 Synthesis of the study L

Starting with a function criterion L, the location of a place 1 in relation to a place 2, we defined after several iterations three production rules. The table 1 is a summary of all iterations performed as well as the production rules defined in this study.

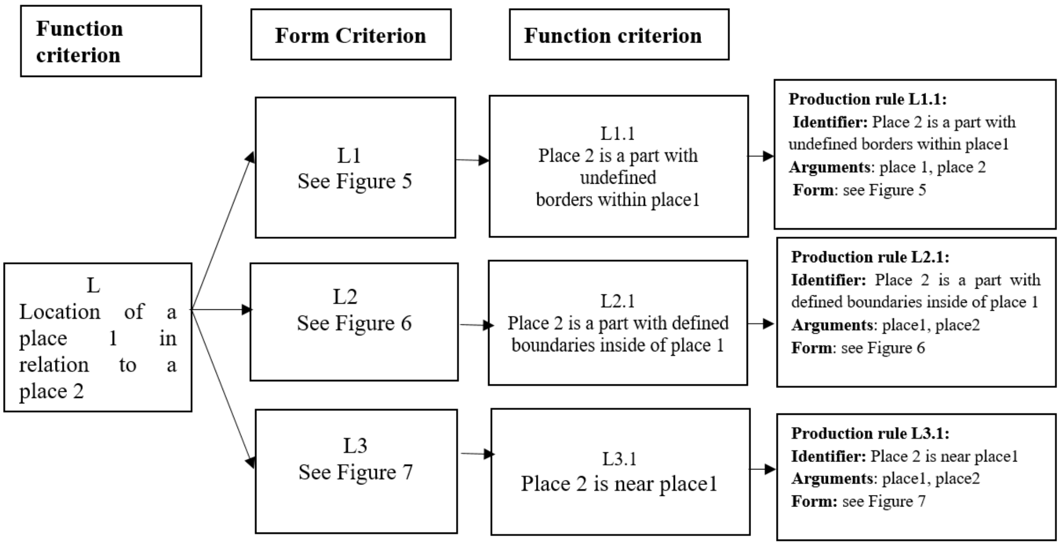


Table 1: synthesis of the study

4. Conclusion

This article has presented the description of some localization structures in LSF. To take into account specificities related to SL, in particular the multilinearity and the use of the signing space, we carried out a study of corpus using a semantic approach. It consists in identifying a systematic link between an observable group of forms and a semantic function. By applying it on the analysis of occurrences of geographical location in the corpus “websourd AFP 2007”, we have identified three production rules relating to the location of a place 1 in relation to a place 2.

It should also be noted that some of the rules defined in this study have been merged with other more global production rules presented in Hadjadj et al. (2018). For example, L.1.1 rule and L.2.1 rule are merged with a rule named “addinfo”. This production rule includes structures whose second item carries additional information to item 1. If we take the example of the production rule L.2.1, the second item gives additional information to item 1 (its geographical location). This important semantic coverage of the rules is interesting for describing different LSF structures, using a reduced number of production rules.

5. Bibliographical References

Cuxac, C. and Sallandre, M. A. (2007). Iconicity and arbitrariness in French Sign Language : Highly Iconic Structures, degenerated iconicity and diagrammatic iconicity. In A. Pizzuto, E. P. Pietrandrea, & R. Simone (eds.), *Verbal and Signed Languages : Comparing Structures, Constructs and Methodologies*. Berlin: Mouton de Gruyter. 13-33.

Filhol, M., Hadjadj, M. N., and Choisier, A. (2014). Nonmanual features: the right to indifference. In Representation and Processing of Sign Languages: Beyond the manual channel, Language resource and evaluation conference (LREC), Reykjavik, Iceland.

Hadjadj, M. N., Filhol, M., and Braffort, A. (2018). Modeling French Sign Language: a proposal for a semantically compositional system. Language resource and evaluation conference (LREC), Miyazaki, Japan.

Huenerfauth, M. (2006). Generating American Sign Language Classifier Predicates for English-to-ASL Machine Translation. PhD thesis, University of Pennsylvania, Philadelphia, PA.

Lejeune, F. (2004). Analyse sémantico-cognitive d’énoncés en langue des signes française pour une génération automatique de séquences gestuelles, Thèse de doctorat, Université Paris 11 Orsay.

Woll, B. (2007). Perspectives on linearity and simultaneity. In M. Vermeerbergen, L. Leeson, & A.Onno (eds.), *Simultaneity in Signed Languages: Form and Function.* Amsterdam: John Benjamins. 103-125.

SLAAASh and the ASL Deaf Communities (or "so many gifs!")

Julie A. Hochgesang
Gallaudet University
800 Florida Ave NE, SLCC 3206, Washington DC 20002
julie.hochgesang@gallaudet.edu

Abstract

The project Sign Language Acquisition, Annotation, Archiving and Sharing (SLAAASh) is a model for working with diverse ASL Deaf communities in all stages of the project. In this presentation, I highlight key steps in achieving this level of collaboration. First, I discuss the importance of sharing work with the community—a key form of reciprocity recognized by Deaf communities. Second, I discuss the importance of reflecting diversity, e.g., ensuring that ASL Signbank actors vary in age, gender, ethnicity, body type, and language experience. Third, I discuss the importance of incorporating feedback from stakeholders and show how the ASL Signbank actors have expressed different views that have impacted our development of the Signbank. Finally, I discuss the crucial component of building substantive community connections and maintaining them long-term. I end by discussing our own efforts to build community connections to date as well as planned future ones.

Keywords: Ethical and legal aspects, experiences in building sign language corpora, language documentation and long-term accessibility for sign language data, ASL

1. Introduction

Sign Language Acquisition, Annotation, Archiving and Sharing (SLAAASh) is a four-year project (2015-2019). Its intent is to systematize annotations and make available to researchers a set of previously recorded longitudinal spontaneous production ASL data with accompanying metadata, annotations, and descriptive analyses. Annotation conventions have been inherited from prior projects as detailed in Chen Pichler, Hochgesang, Lill-Martin & Quadros (2010) but revised based on current best practices for sign language documentation (Chen Pichler, Hochgesang, Lillo-Martin, 2015). With the help of the technical team that built the NGT Signbank, a Signbank has been developed for ASL (Hochgesang, Crasborn, and Lillo-Martin, 2017). To date, almost 40 percent of the acquisition video data has been annotated, community input has been collected on reconsenting and sharing protocol (Chen Pichler et al., 2016; 2016), and over 2700 entries have been added to the ASL Signbank.

As evidenced by the reconsenting and sharing protocol already reported (Chen Pichler et al., 2016; 2016), SLAAASh believes it is essential to work with the signed language communities throughout all stages of the project, continually reporting progress and seeking input (e.g., Harris, Holmes & Mertens, 2009). This paper presents the ways that the SLAAASh project has worked with the ASL Deaf communities. First, I discuss the importance of sharing work with the community—a key form of reciprocity called for by Deaf communities. Second, I discuss the importance of reflecting the diversity found in those communities. For example, ensuring that ASL Signbank actors vary in age, gender, ethnicity, body type, and language experience. Third, I discuss the importance of incorporating feedback from stakeholders and show how the ASL Signbank actors, in expressing their individual views, have impacted our development of the Signbank. Finally, I discuss the crucial component of building substantive community connections and maintaining them long-term. I will end by discussing our own efforts to build community connections to date as well as planned future ones.

2. How We Share the Work We Do With the Community

Through the usual academic channels of workshops, conferences, and papers, SLAAASh has started to report on its activities (e.g., Hochgesang et al., 2017). These channels, however, are not as accessible to Deaf communities. Therefore, SLAAASh has used Twitter (@ASLSLAASH) to steadily stream project news, most particularly updates to the ASL Signbank.

Pictures and animated GIFs are used to introduce ASL Signbank actors (Figure 1), ASL signs and even the name sign of the Signbank itself. Providing information in written English alone does not provide enough access to the community. Putting aside the varying levels of literacy skills of Deaf people, the language we are looking at – ASL - simply does not have a conventionalized written system that can be used to fully and adequately represent it. Since Twitter allows for mixed media, different digital tools can and should be used to share information – typed text, pictures, videos, and animated GIFs. While videos and animated GIFs are ideal for transmitting signed messages, they cannot be used solely because they are not searchable without written text (as any person working with signed corpora knows).

https://twitter.com/aslslaash/status/757641765109149696
Figure 1. Example of tweet introducing one of the actors for the ASL Signbank

The name sign for the ASL Signbank even evolved because of the GIF alone. At each filming for the Signbank or presentation about the Signbank, GIFs of the ASL Signbank name sign were taken or shared. Discussions always followed, thus leading to an evolution of the name sign itself.

Figure 2. Examples of GIFs for the "ASL Signbank" name sign

Name signs are usually designated by members of Deaf communities. While a name sign could be decided by a single person, it also can evolve through discussion and negotiation with other members. At a guest lecture for the National Association of the Deaf (2017), I introduced the ASL Signbank to the staff and demonstrated the name sign that was current then – as Figure 2 shows it was a combination of SIGN and FS(bank). The NAD members were not keen on having fingerspelling as part of the name sign so they offered their own version seen in Figure 3 below. They chose another version of the word "to sign". The one in Figure 2 is a more neutral and noun-like version referring to the general ability to sign or the modality itself; the one in Figure 3 is more related to the sense of using a word (or sign) itself and is often used to identify people who can sign fluently or are "closer to the Deaf community" than others.

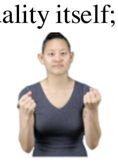


Figure 3. NAD suggestion for ASL Signbank name sign

Then the NAD staff chose a depicting sign to represent "bank" in a way that refers to putting things in a repository rather than the financial institutions as signified by the English word "bank". When I shared this name sign with yet other audiences, they appeared to approve on the artistry or novelty of this expression and for using it to describe the ASL Signbank. But for referring to the ASL Signbank (or using it as a name), there was a preference for a more "streamlined" sign as a name sign rather than a phrase. The name sign suggested by NAD was modified so that the first word ("to sign") became a one-handed version (although the sign usually resists weak hand drop) and combined with a one-handed depicting sign that indicates a list. This new name sign, shown in Figure 4, can be reproduced as two-handed and moved in a way to indicate pride or just one-handed for plain reference.

Figure 4. The latest name sign for the ASL Signbank

The new name sign has the added (and coincidental) bonus of resembling the letters  ("s") and  ("b") in the manual alphabet used in ASL.

This ongoing dialogue with the community members and the evolution of the name sign for the ASL Signbank exemplify the first two principles of the Sign Language Communities' Terms of Reference (SLCTR) by Harris, Holmes and Mertens (2009) :"(t)he authority for the construction of meanings and knowledge within the Sign Language community rests with the community's members" and the second principle, "(i)nvestigators should acknowledge that Sign Language community members have the right to have those things that they value to be fully considered in all interactions" (115).

Sharing research done on signed languages is essential. Deaf communities appreciate (if not often demand) reciprocity (e.g., Harris et al., 2009, 115). If the Deaf communities contribute a part of their lives by demonstrating how they use their language, then the researchers need to reciprocate by sharing the work built on this language use in an accessible manner. Academic products like articles or conference proceedings are not as accessible as social media channels (e.g., Twitter or Facebook). Sharing the work done by SLAAASh via Twitter made it possible to enter into a rich and ongoing exchange about how to refer to a lexical database. While we have the academic spoken/written name of the product, we now also have a name sign, something that is valued by American Deaf communities.

3. Diversity in the ASL Signbank

Clearly every community is diverse. While we can identify features that characterize a certain community, this does not mean the communities are homogenous (e.g., Harris et al., 2009). Although there may be necessary generalizations, any work with language needs to reflect the diversity of the communities. For example, we say that

ASL is the sign language of the Deaf community in America. But the truth is there are multiple varieties for the multiple Deaf communities in America. The lines are not always easily drawn. Nor are the communities neatly mapped onto the different varieties. Unsurprisingly so because identities are intersectional and signers can choose to use specific variants depending on who they are interacting with and why – what Eckert calls "speaker agency" (Eckert, 2008). Any language documentation project is then ethically obligated to reflect the authentic diversity of the researched language communities. Perfect representation (for anything), for countless reasons, is impossible to attain but the ongoing, transparent and reflective attempt to recognize and represent different experiences in itself is valuable. I discuss two ways we do so with SLAAASh – ensuring diversity of signers in the ASL Signbank and representing any and all ASL signs that are in the corpora using the ASL Signbank.

3.1 Diversity of the ASL Signbank Actors

The source of the lexical items included in the ASL Signbank, of course, comes from the primary data of the SLAAASh project (child ASL acquisition videos as well as other kinds of videos associated with the other projects that use the ASL Signbank, e.g., Philadelphia Signs Project). The original videos, for both confidentiality and quality issues, cannot be edited and re-used as ASL Signbank videos to represent the lexical items themselves. Thus it is necessary to hire actors to produce clear, isolated and unmodified (e.g. for grammatical aspect) forms of the signs that can be used in representative movies in the ASL Signbank (much like how we have the basic form of a word as headwords or "lemmas" in dictionaries).

Our actors are native or early users of ASL (early meaning preferably the user acquired ASL before the age of four). The current lineup of ASL Signbank actors (shown in Figure 5) is approaching representation of the different Deaf communities in America who use ASL.

Because we do not have reliable survey demographic data, it is not possible to discuss whether the ASL Signbank actor demographics are proportionate to the American Deaf communities. Since the ASL Signbank itself is not a corpus, that is not really a concern in terms of corpus representativeness. However, because ASL Signbank is a representation of the ASL lexicon, the actors themselves do need to be diverse because the American Deaf communities are. Thus, our signers vary in age, gender, ethnicity[1] and body type (as is demonstrated in Figure 5).

They also vary in kinds of "language experience", which, in this paper, means the age of acquisition along with the type of language input.

Figure 5. Current lineup of ASL Signbank actors

While SLAAASh requires that the ASL Signbank actors have had acquired ASL before the age of four, there are no other requirement because we recognize that the language experiences of the American Deaf communities are varied – some are raised in hearing families, some are hearing themselves with Deaf parents, and et cetera. They all make up the American Deaf community experience.

3.2 Diversity of Lexical Items in the ASL Signbank

Since the ASL Signbank is a research tool, specifically a lexical database that can be linked directly to annotation software, it needs to be able to handle whatever comes up in the data. That means sometimes including forms that some members of the American Deaf communities may not consider part of their ASL use. One simple example is regional variants, e.g., the *soda/pop/etc* variants in American English. The different ASL variants for "birthday" in Figure 6 is an example of this kind of regional variation.

Figure 6. ASL regional variants for "birthday"

In addition to including all regional variants that occur in the corpora using the ASL Signbank, there will be a feedback function when the ASL Signbank is made public where users can contribute their own regional variants.

[1] On a personal note, there was something my student said a few years back. "It means a lot to see someone that looks like me up there on the screen. I don't usually see that." She was a person of color and felt under-represented in presentations and publications about signed languages. She was struck by my class presentation which included people of varying ethnic backgrounds. It made me even more dedicated to ensure a wide representation of users in my own work.

Another example of variation in ASL results from the influence of spoken languages or invented manual codes used in education to represent these spoken languages. As shown in Figure 7, the sign for "the" is an example of this influence from a manual code intended to help written English in American Deaf education.

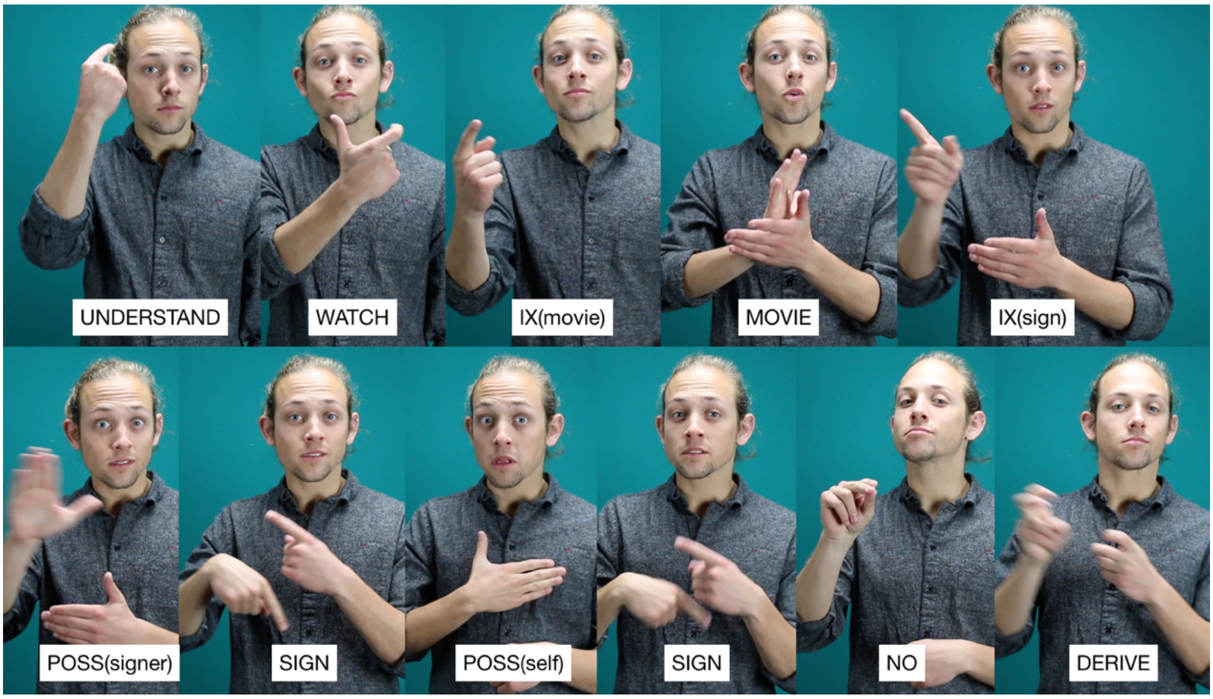
Figure 7. Sign for "the"

The use of these forms can be quite controversial in the American Deaf communities and invoke discussions about which signs are "real" or "right". Being a usage-based and descriptive research tool, the ASL Signbank includes all variants that arise in the data. But being mindful of the uncomfortable issues they can incite in the American Deaf communities, the SLAAASh research project adds information to the ASL Signbank to reflect current language attitudes (see Hochgesang, Crasborn, and Lillo-Martin this volume for more on the ASL Signbank design). For example, in the *morphosyntax* section, the "derivation history" field can be used to categorize signs as being initialized (i.e., signed with the handshape that represents the first letter of the ambient spoken language) or fingerspelled. We also will add a "usage" section in which we can add memos reporting observations shared by the community, like usage notes in dictionaries ("polite", "vulgar slang", "offensive", "old-fashioned"). This strategy allows us to navigate language attitudes while accurately representing the data.

4. Views of ASL Signbank Actors Feeding Back Into Our Work

It is also important to consider ASL users' attitudes toward various signs while accurately representing the data (e.g., Harris, Holmes and Mertens, 2009). One way the SLAAASh project has been able to do this is through dialogues with the ASL Signbank actors during filming sessions.

Some of the ASL Signbank actors sometimes expressed discomfort with producing certain variants during filming. Perhaps they just did not know these variants and they felt too unfamiliar for their hands. Or others were current ASL teachers who did not want to film certain signs that could be used against them in their professional work. Yet other signs were considered to be offensive or taboo. The actors either opted out of filming those (with our full support) or filmed them with the understanding that a "disclaimer" would be posted on the website specifying that these signs are not necessarily the typical productions of the actors



"When you watch the videos, remember that these signs may not be the signs that the actors actually use. They are just re-producing what appeared in the primary data."

Figure 8. Example of disclaimer that will be displayed on the ASL Signbank website

themselves. This is just one example of the discussions that arose from the ASL Signbank filming.

We intend to include brief videos on the ASL Signbank website that has resulted from these discussions - that the signers are actors, that the database is not a dictionary but a research tool and so on. For example, Figure 8 is a series of stills taken from a brief video explaining that signs produced by the actors in the ASL Signbank may not be the variants they actually use.

5. Community Connections

Finally, there are several community connections to the project, both ongoing and ones planned for the future. For example, we will give presentations to the community with information about the ASL Signbank and how it can be used for personal interest, teaching ASL, Deaf education, and other extended uses. These presentations will be made available online as well as face-to-face, in order to reach a wider audience (and they will be announced on @ASLSLAASH). Members of the ASL community will be included in the Advisory Board that will review applications to access the acquisition data, and they will serve as advisors to the online repository for the acquisition data. Other projects are already making use of the ASL SignBank for their own purposes, including Philadelphia Signs, multiple departments at Gallaudet (e.g., Linguistics, Department of Interpretation and Translation, ASL and Deaf Studies, and Education), sign language researchers from Eastern Kentucky University, and Boston University, and some early educators of Deaf children. Access to the SignBank will be under a Creative Commons license (https://aslsignbank.haskins.yale.edu//about/copyright/) through which users will be encouraged to share their own work making use of the ASL Signbank, to further enrich American Deaf communities.

6. Conclusion

As already mentioned throughout the paper, I refer to the SLCTR developed by Harris et al. (2009) throughout my own work. Harris et al. (ibid) discuss the ethics of research, particularly with un- or under-represented groups and specifically with signed language communities. They also propose a set of "culturally appropriate research guidelines" intended to accord respect and show sensitivity towards the studied group's culture. They are not the only resource available to the signed language researcher who wishes to consider ethical aspects of working with signed language communities. *Working Together – Manual for Sign Language Work within Development Cooperation* (http://www.slwmanual.info) presents guidelines in both English and International Sign. Also, our own work from the SLAAASh project described here stands as a model for ethical considerations when working with signed language communities. Whatever the resource, it is necessary to continually engage and involve the relevant Deaf communities however possible. The languages we research come from their own hands and lives.

7. Acknowledgements

The research reported here was supported in part by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award number R01DC013578 and award number R01DC000183. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Amelia Becker, Terra Edwards, and Diane Lillo-Martin for their beneficial input/additions to this paper.

8. Bibliographical References

Chen Pichler, D., Hochgesang, J.A., Lillo-Martin, D., & de Quadros, R. (2010). Conventions for sign and speech transcription in child bimodal bilingual corpora. Languages, Interaction and Acquisition, 1(1), 11-40.

Chen Pichler, D., Hochgesang, J.A., Simons, D. & Lillo-Martin, D. (2016). Community Input on Re-consenting for Data Sharing. Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining. 29-34. LREC, Portorož, Slovenia, May 28 2016.

Chen Pichler, D., Hochgesang, J.A., & Lillo-Martin, D. (2015, March). BiBiBi Project ASL Annotation Conventions. Poster presented at "Digging into Signs Workshop: Developing Annotation Standards for Signed Language Corpora". University College London, London, United Kingdom (March 30-31, 2015).

Chen Pichler, D., Hochgesang, J.A., Simons, D., & Lillo-Martin, D. (2016, Jan). Reconsenting for Data Sharing. Poster presented at the Theoretical Issues in Sign Language Research 12 Meeting, LaTrobe University, Melbourne, Australia (January 4-7, 2016).

Crasborn, O. & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood & E. Thoutenhoofd (Eds.), Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages. ELRA, Paris, pp. 39-43.

Eckert, P. (2008). Variation and the indexical field. Journal of Sociolinguistics 12/4, 2008: 453–476

Harris, R., Holmes, H.M., & Mertens, D.M. (2009). Research ethics in sign language communities. Sign Language Studies, 9 (2), 104-131.

Hochgesang, J.A. (2017, May). Ethics and my research. Invited presentation for NAD Board "Monthly Training Series". Silver Springs, Maryland.

Hochgesang, J.A., Crasborn, O., & Lillo-Martin, D. (2017) ASL Signbank. New Haven, CT: Haskins Lab, Yale University. https://aslSignbank.haskins.yale.edu/

Hochgesang, J.A, Crasborn, O., & Lillo-Martin, D. (2017, July 24). Building the American Sign Language (ASL) SignBank on the Shoulders of Giants. Poster presented at the workshop "Corpus-based approaches to sign language linguistics: Into the second decade", 9th International Corpus Linguistics Conference, University of Birmingham, UK, July 24-28, 2017.

9. Language Resource References

ASL Signbank (2018). Lexical database for American Sign Language. URL: http://aslsignbank.haskins.yale.edu/.

10. Appendix – Sign Language Communities' Terms of Reference (SLCTR) Principles

The six principles of the Sign Language Communities' Terms of Reference (SLCTR):

1. *The authority for the construction of meanings and knowledge within the Sign Language community rests with the community's members.*
2. *Investigators should acknowledge that Sign Language community members have the right to have those things that they value to be fully considered in all interactions.*
3. *Investigators should take into account the worldviews of the Sign Language community in all negotiations or dealings that impact on the community's members.*
4. *In the application of Sign Language communities' terms of reference, investigators should recognize the diverse experiences, understandings, and way of life (in sign language societies) that reflect their contemporary cultures.*
5. *Investigators should ensure that the views and perceptions of the critical reference group (the sign language group) is reflected in any process of validating and evaluating the extent to which Sign Language communities' terms of reference have been taken into account.*
6. *Investigators should negotiate within and among sign language groups to establish appropriate processes to consider and determine the criteria for deciding how to meet cultural imperatives, social needs, and priorities.*

(Harris, Holmes, Mertens 2009, 115).

11. Appendix – Author's Positionality

I am a Deaf American woman born to a hearing white family in the late 1970's. My parents learned a variety of American Sign Language with me – one that was influenced by the popular belief then that a manual code (Signed Exact English) should be used to facilitate the learning of English. I attended mainstreamed schools from kindergarten throughout high school although the type of program and services varied –a " total-communication" self-contained (i.e., only with deaf students requiring the same kind of services) program with other students (mostly hard of hearing or oral) ; as a single student with an interpreter in all-hearing classes ; in a mainstreamed program with other Deaf students who used American Sign Language but usually not taking self-contained classes. I wanted to transfer to a Deaf residential school but was advised not to because I would have had to move up two academic grades. Through different interactive opportunities (Deaf camps, Deaf community theatre, and Deaf social events), I was able to interact daily with the Deaf communities in Northern Illinois. By the time I was in high school, I started to actively reject speech therapy and manual codes for English – taking pride in my use of American Sign Language (although I remained a passionate reader of written English literature). For college, I went to California State University at Northridge – at that time it had a large Deaf program – about 200 Deaf students. After graduating, I joined the Peace Corps and lived in Kenya where I taught at a Deaf school for two years. I learned their signed language – Kenyan Sign Language (which bears historical influence from American Sign Language, British Sign Language, Swedish Sign Language, and possibly others). Upon return to the States, I attended Gallaudet for graduate school in linguistics and got my PhD. I also taught in the ASL and Deaf Studies program. After two years I transferred to the linguistics department where I am now an assistant professor. During my time at Gallaudet, I also married a hearing black man (I consider myself an ally of the LGBT community) and have two multi-racial boys. I also have had several language documentation experiences with multiple Deaf communities – both American and international. *All of these experiences have shaped me and instilled in me a deep respect for diversity and individual experiences.*

https://twitter.com/jahochcam/status/828415245421047812

Building the ASL Signbank: Lemmatization Principles for ASL

Julie A. Hochgesang[1], Onno Crasborn[2], Diane Lillo-Martin[3]
[1]Gallaudet University, [2]Centre for Language Studies, Radboud University,
[3]University of Connecticut & Haskins Laboratories
800 Florida Ave NE, SLCC 3206, Washington DC 20002, USA; PO Box 9103, 6500HD Nijmegen, Netherlands;
Department of Linguistics, 365 Fairfield Way, Unit 1145, Storrs, CT 06269-1145, USA
julie.hochgesang@gallaudet.edu, o.crasborn@let.ru.nl, diane.lillo-martin@uconn.edu

Abstract

Following the example of other sign language researchers, we are creating a Signbank, a usage-based lexical database, to maintain consistent and systematic annotation information for American Sign Language (ASL). This tool, which will be available to the public, is currently being used in conjunction with an on-going effort to prepare corpora of sign language acquisition to share with the research community. This paper will briefly report on the development of the ASL Signbank, focusing on the adopted lemmatization principles. Lemmatization of ASL signs has never been done on a scale like this before - one that has been continually refreshed by actual usage data.

Keywords: lexical database, lemmatization, ASL, signbank

1. Introduction

Signbanks, usage-based lexical databases, have been created for several signed languages (Auslan, Johnston 2001; British Sign Language, Fenlon et al. 2014; Sign Language of the Netherlands, Crasborn et al. 2016; Finnish Sign Language, Salonen et al. 2016). Given the lack of conventionalized writing systems for signed languages, a best practice for annotating is to use ID glosses (Johnston 2001), unique gloss identifiers of signs. To keep an organized database of ID glosses and the signs they represent, they are added to the Signbank as the signs are observed in the primary data while annotating. Subsequently, the signs are organized in the database using lemmatization principles. We are creating a Signbank for American Sign Language (ASL), currently used in conjunction with SLAAASh, an ongoing effort to prepare corpora of sign language acquisition to share with the research community. The Signbank itself is to be made available to the public for general use. This paper will briefly report on the development of the ASL Signbank (which now has 2600+ entries), focusing on the lemmatization principles adopted from Fenlon et al. (2015), as well as on the workflow we have implemented for creation and maintenance of ID glosses We also touch upon how we use ASL Signbank for research.

2. SLAAASh and ASL Signbank

The Sign Language Acquisition, Annotation, Archiving and Sharing (SLAAASh) project is working with a digitized video corpus of Deaf children's use of ASL, collected as spontaneous production data from 4 Deaf children of Deaf parents, ages 1;04-4;01 (Lillo-Martin & Chen Pichler 2008). We are currently annotating the primary data systematically using our ID glosses and annotation conventions. We also are engaging re-consenting protocols (Chen Pichler et al. 2016) to document permission from the children (who are now adults) and others in the recordings to share their data for research purposes. And lastly, we are also working with others to develop a web-based platform for sign language data sharing. Taken together these activities create an annotation, archiving and sharing infrastructure that can be used by other research projects also studying ASL, exponentially increasing the availability of usage-based observations of signs for research.

The entries in the ASL Signbank are produced and coded by our Signbank team headquartered at Gallaudet University. This system is allowing us to organize the ID glosses we have been developing over many years throughout the various incarnations of our projects (Chen Pichler et al 2010; Chen Pichler et al 2015). At earlier stages of this process we used homegrown efforts to organize ID glosses (single folder on a single user's computer, shared Google Drive account, shared Dropbox account). Soon we discovered that these attempts were inefficient and that we needed to turn to a lexical database solution like a Signbank.

The ASL Signbank software is modelled on the NGT Signbank, which in turn is based on the Auslan Signbank software (Cassidy et al. 2018). The software is available for developers under a public license at http://github.com/Signbank/Global-Signbank/. The ASL Signbank infrastructure has been developed and maintained by Radboud University, but it is hosted by Haskins Laboratories and Yale University in the US. In addition to organization and access, an advantage of the ASL Signbank is the availability of direct linking to the ELAN annotation software (Crasborn et al. 2016), as part of version 5.0 (released October 2017).

2.1 ASL-LEX

The ASL Signbank team is collaborating with the team building ASL-LEX (Caselli et al 2016), a publicly-available database which includes subjective frequency and iconicity judgments as well as phonological information on 1,000 signs (and more to come). Our collaboration involves sharing ID glosses, so that signs that are common across the databases can be easily accessed, as well as phonological information, so that the signs have consistent coding.

We are building up both projects simultaneously, coordinating new entries as possible while accommodating the distinct project requirements. The goals of the ASL Signbank and ASL-LEX are somewhat different: the Signbank is based on usage data (e.g., ID glosses for signs are created as they occur in our child acquisition data, as well as in the data from other research projects that use our ID glosses), while the ASL-LEX project was designed to include elicited signs in order to represent the full range from high to low frequency and high to low iconicity, for use in psycholinguistic experiments. Despite these different goals, the projects are mutually reinforcing. Our projects are linked together by the alignment of glosses (we use the same lemmas, although we may differ in annotation ID glosses, described in sections 3 and 4); shared phonological coding (using a simplified version of the Prosodic Model (Brentari 1998)); shared iconicity ratings (subjective iconicity ratings as well as iconicity categorization); and shared lexical properties (e.g., identification of lexical class). Eventually, the actual frequency data from our corpora (in child signing and child-directed signing) will help to tie the projects even closer together.

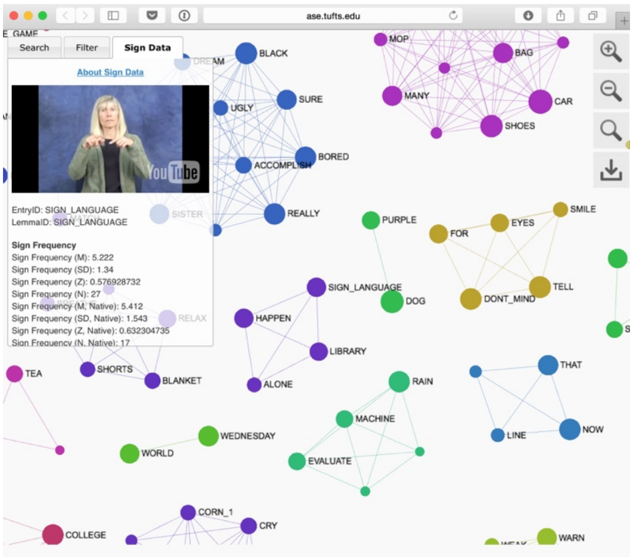


Figure 1. ASL-LEX, visualization view of phonological neighborhoods of some ASL signs

3. Overall description of signs

As with other Signbanks, our goal is to create an open-access lexical database of ASL signs with their ID glosses (Johnston 2001) to facilitate consistent and systematic annotation of sign usage in multiple data sets. Along with a movie and image of the sign and its ID gloss, each entry has information about the sign's formational components, its grammatical characteristics, and usage information. The categories of information about each sign used in the ASL Signbank have been derived from all prior Signbanks and prior annotation conventions (Chen Pichler et al 2015). As discussed briefly in 2.1, we also consider the data categories used by ASL-LEX. To illustrate, Figure 2 shows a record from the ASL Signbank for AGAIN.

Using our lemmatization principles (outlined in section 4), we have a lemma ID gloss for the sign as well as an annotation ID gloss which will be slightly different if the sign has phonological variants (which occurs when forms share all phonological features except for one or two). We enter "translation equivalents" (keywords) to facilitate the search for each sign and to represent the meaning of the sign. These can also be used in ELAN when the ASL Signbank is used as an external controlled vocabulary (ECV). This reduces the need for annotators to memorize ID glosses. The dialect field allows us specify any US region. The field "semantic field" is used to group together sets of signs that refer to the same subject. We inherited the fields from the NGT Signbank for the *morphology* section, allowing us to describe the make-up of compounds (See Crasborn et al. 2016). In the *phonology* section (coded in conjunction with ASL-LEX), we identify handedness, major location, minor location (beginning and final), dominant hand selected fingers and flexion as well as any abduction or flexion change, nondominant handshape and path movement. For the *morphosyntax* section, we identify the word or lexical class of the sign as well as its derivation history (lexicalized through fingerspelling, compounding, borrowing, et cetera), and type of iconicity. *Relations to other signs* allows us to connect ID glosses with homonyms, synonyms, variants, antonyms, hyponyms, hypernyms, and so on. *Relations to foreign signs* tracks any known connection with borrowed signs from other signed languages. *Frequency* will mark how many times the sign occurs in our corpus as well as the number of signers in the corpus who use that sign. *Publication status* and *notes* allow us to add metadata about the entry itself, especially useful for maintenance of ID glosses if they need to be changed.

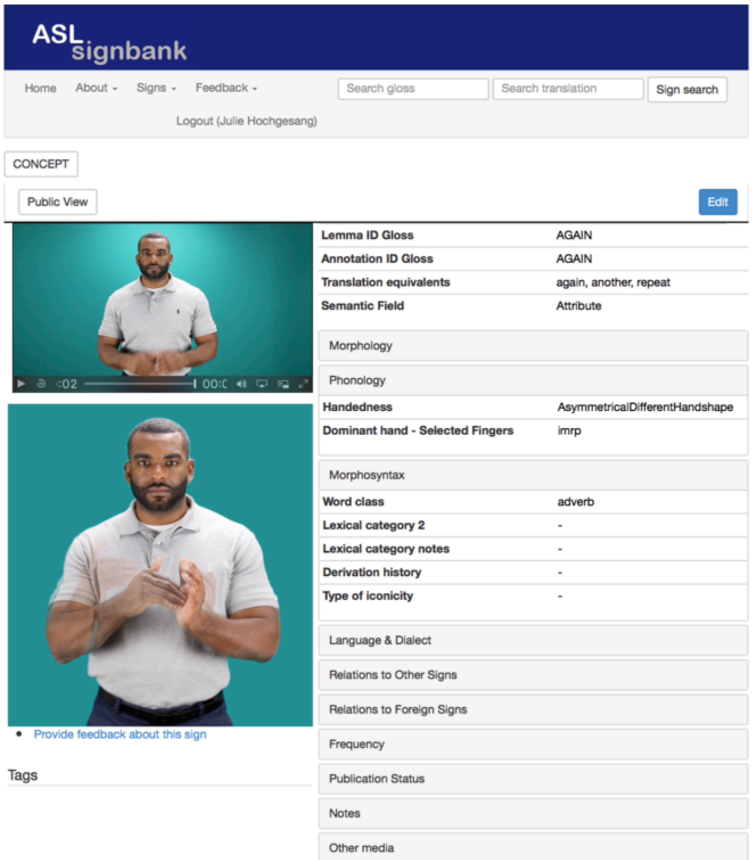


Figure 2. ASL Signbank entry for AGAIN

4. Lemmatization of ASL signs

The data in the SLAAASh corpus comes from four Deaf children, their Deaf parents, and others interacting with them. Clearly, this cannot be considered a representative selection of ASL signers or even of ASL acquirers. In this way, the SLAAASh corpus-building may be seen as different from current sign language corpora projects like the BSL Corpus project. Nonetheless, our treatment of signs is the same – each citation form gets its own gloss. We did not pre-determine a list of signs to be included (as dictionaries might do), but assign ID glosses as we come across the signs in the primary video data. Initial assignment of ID glosses did not follow lemmatization principles by determining which forms are related to which lexemes; the only real rule we had was to give each sign form a different ID gloss. After enough entries began to accumulate, we were able to modify the organization and assignment of ID glosses on the basis of lemmatization principles described in Fenlon et al (2015). To our advantage, the connection between Signbanks and ELAN made possible in recent releases (including an external controlled vocabulary generated by a Signbank, and a Signbank Lexicon Service in ELAN) permits changes in annotation glosses recorded in ASL Signbank to be promulgated throughout the annotations in our corpus.

We generally follow the same principles as laid out in Fenlon et al (2015):

> *…we consider the citation form to be the lemma (i.e. the unmodified form of a given sign is used here as the headword of a lexeme)…The ID gloss is a unique English-based translation used primarily as an annotation tag in the corpus for all occurrences of that lexeme regardless of how it might be modified. It is important to note that the choice of the English word as an ID gloss for a particular lexeme is not meant to indicate the sign's core meaning or grammatical function. It is merely a label to uniquely identify each lexeme, to be used in annotation of sign language data, in lieu of any standardised orthography for the language. [However] (for) the purposes of annotation… it is much more useful to use ID glosses that have some meaningful connection to the lexeme, e.g. via one of the translation equivalents, since annotation is done by typing in the ID gloss" (176).*

The lemmatization principles are simple at first glance: if the meaning and the form of two entries are different, they are different lemmas and get different ID glosses; if the meaning is similar and there are one or two phonological differences, they are under the same lemma and get the same core annotation ID glosses, with the difference indicated by lowercase tags that identify the particular formational aspect responsible. For example, the ASL signs for "believe" and "precious" (Figure 3) are clearly different forms. One is two-handed, the other is one-handed.

Figure 3: ASL signs for "believe" and "precious" with their ID glosses

There are different locations, handshapes, path, and other formational features characterizing each sign. Also the meanings are, of course, quite different. So with that, they are deemed different lemmas and accordingly get unique glosses.

In Figure 4, it can be seen that two forms for 'believe' are similar but clearly different in at least one aspect, specifically the initial handshape for the dominant or strong hand (for the first and for the second). Their meaning, however, is the same. Given that, we treat them as phonological variants linked to the same lemma but with unique annotation ID glosses (marked by the lowercase tags, i.e., "b" and "ix").

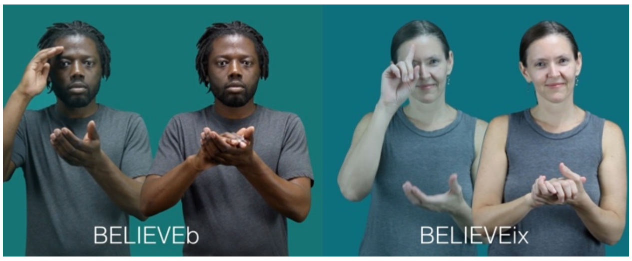


Figure 4: ASL variants for "believe" with their ID glosses

So, signs with phonological variants (e.g., BELIEVEb and BELIEVEix) are annotated with distinct ID glosses in our ELAN files, which are linked to ASL Signbank (and connected under the same lemma). The signs in Figure 5 are examples of signs under different lemmas. They are synonyms but have clearly distinct phonological forms.

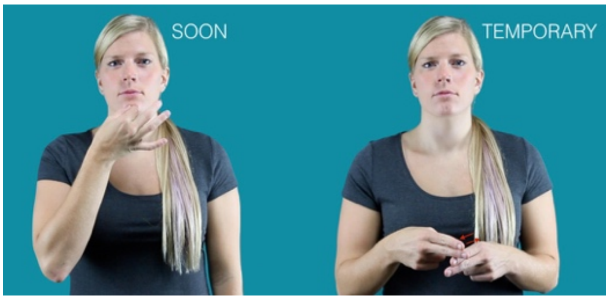


Figure 5: ASL signs for "soon" and "temporary" with their ID glosses

The signs for "soon" and "temporary" receive unique ID glosses but are marked as "related" in the ASL Signbank.

In practice, this adherence to lemmatization has been remarkably difficult but instructive. For instance, the ASL sign for "equal" (Figure 6) can have a single set of movements (bent hands move together to touch) or repeated sets.

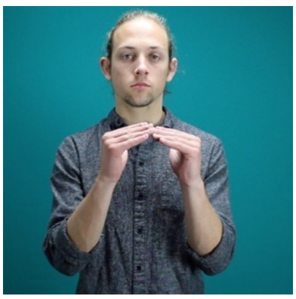


Figure 6: ASL sign for "equal"

The "equal" forms will vary based on syntactic placement as well as intended meaning. Are they of a single lemma with one of the forms a modification of the lemma? Or are they two separate lemmas with conventionalized separate meanings? We have tentatively left this example as a single lemma with one annotation ID gloss. We will revisit it once we have a sufficient number of examples in the corpora that use the ASL Signbank.
On the other hand, the ASL signs for "show" and "example" are produced similarly but one ("show") has one set of movements and the other ("example") has shorter and repeated movements (Figure 7). With this slightly different phonological form and their conventionalized different meanings, they are separate lemmas.

Figure 7: ASL signs for "show" and "example".

These lemmatization decisions are made both by observing how these signs behave in the dataset and how they are understood by the researchers. Regular lab meetings are held to discuss sign lemmas. Frequently the discussion involves producing the sign in various modifications. For example, if the two signs are verbs, they can be modified to reflect grammatical aspect. If they are changed in the same way, they are deemed the same lemma. For example, these two forms in Figure 8 appear to be the same at first glance but they are used in different contexts and cannot be changed in the same way when modified for aspect.

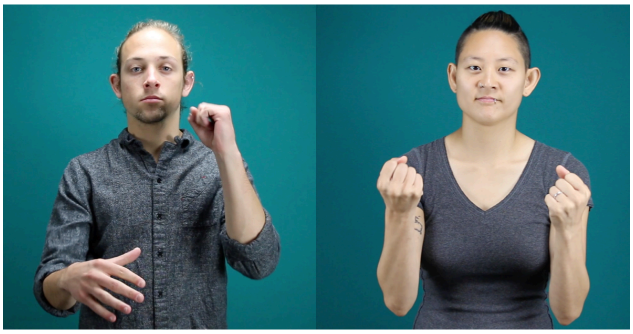
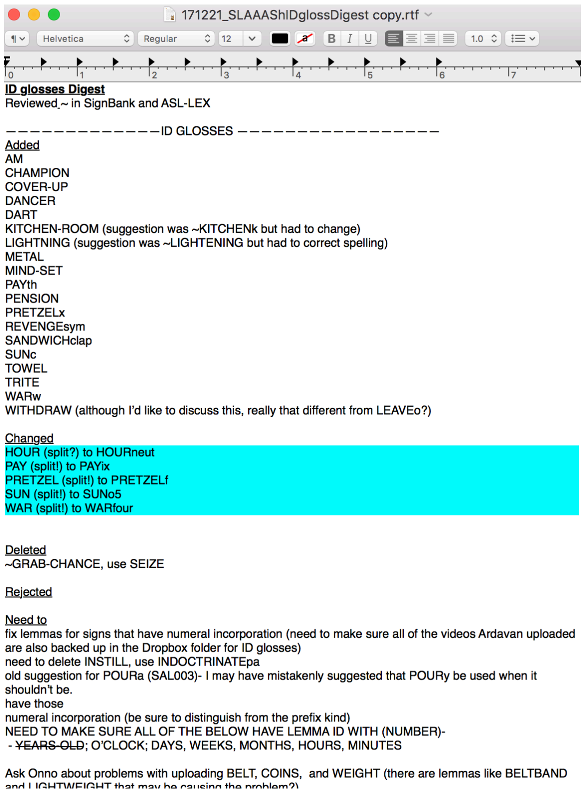


Figure 8. ASL forms basically meaning "to sign" but are different lemmas because they are used differently

For signs that may not be as lexically fixed but are specific types of signs, we use regular codes to annotate them, e.g., DS for "depicting sign", NS for "name sign", IX for "index" or pointing sign. Specific referents are added to related tiers in the annotation files. (See SLAAASh annotation conventions for more, (Hochgesang (2015)).

5. Workflow of creation and maintenance of ID glosses

The maintenance of the ID glosses is under the supervision of one person, currently Julie Hochgesang, on the SLAAASh research team. All of the annotators for SLAAASh and other research projects who use the ASL Signbank are required to follow a specific protocol for suggesting an ID gloss when the sign they need to annotate is not in the Signbank. After ensuring that they have exhausted all possibilities by searching translation equivalents on ASL Signbank, they then propose an ID gloss using their understanding of the ASL Signbank lemmatization principles. Since the SLAAASh eafs are linked to the ASL Signbank ECV, they must force the annotation field to escape the list in order to enter new entries. They prefix their suggestion with ~, e.g., ~PROPOSED-NEW-IDGLOSS. They then take a video of themselves producing the sign and upload this to the ASL Signbank. They click "proposed new sign" and add the tag "proposed new ID gloss needs review". These three steps are a triple safeguard against errors in the annotation files and accidentally adding them as permanent additions to the ASL Signbank.
The ID gloss supervisor then reviews the suggestions and ensures that the additions are not duplicates of already existing signs. Lemmatization principles as outlined in section 4 are applied. The sign is then marked as approved and tagged to be refilmed. Native/early signers are hired to produce signs which are then published in ASL Signbank. ID gloss digests (shown in Figure 9) are reports of additions, deletions, changes, ongoing issues and are shared with the entire SLAAASh project team as well as others who are registered users of the ASL Signbank.



171221_SLAAAShIDglossDigest copy.rtf

ID glosses Digest
Reviewed ~ in SignBank and ASL-LEX

——————————ID GLOSSES ——————————
Added
AM
CHAMPION
COVER-UP
DANCER
DART
KITCHEN-ROOM (suggestion was ~KITCHENk but had to change)
LIGHTNING (suggestion was ~LIGHTENING but had to correct spelling)
METAL
MIND-SET
PAYth
PENSION
PRETZELx
REVENGEsym
SANDWICHclap
SUNc
TOWEL
TRITE
WARw
WITHDRAW (although I'd like to discuss this, really that different from LEAVEo?)

Changed
HOUR (split?) to HOURneut
PAY (split!) to PAYix
PRETZEL (split!) to PRETZELf
SUN (split!) to SUNo5
WAR (split!) to WARfour

Deleted
~GRAB-CHANCE, use SEIZE

Rejected

Need to
fix lemmas for signs that have numeral incorporation (need to make sure all of the videos Ardavan uploaded are also backed up in the Dropbox folder for ID glosses)
need to delete INSTILL, use INDOCTRINATEpa
old suggestion for POURa (SAL003)- I may have mistakenly suggested that POURy be used when it shouldn't be.
have those
numeral incorporation (be sure to distinguish from the prefix kind)
NEED TO MAKE SURE ALL OF THE BELOW HAVE LEMMA ID WITH (NUMBER)-
- ~~YEARS-OLD~~; O'CLOCK; DAYS, WEEKS, MONTHS, HOURS, MINUTES

Ask Onno about problems with uploading BELT, COINS, and WEIGHT (there are lemmas like BELTBAND and LIGHTWEIGHT that may be causing the problem?)

Figure 9: Screenshot of ID Gloss Digest

6. Use of ASL Signbank for Research

The ASL Signbank contributes to research in two broad ways. First, there are research projects that make use of the data in the Signbank itself, along with the connections to ASL-LEX. Second, there is research that is enabled by the use of Signbank in annotation of sign language data such as the SLAAASh project.

Because the Signbank includes information about the form of each sign as well as morpho-syntactic and other information, it is possible to conduct analyses based on the signs in the Signbank that would typically exceed the number of examples based on other methods. For example, analyses of the frequency of occurrence of specific phonological elements (e.g., selected fingers) can readily be made to test previous claims about markedness. Another example is the occurrence of forms that violate Battison's (1974) symmetry and dominance constraints. Although our original coding system assumed these constraints would hold, we discovered a (relatively small) number of signs that violate the constraints, and adjusted the phonological coding options accordingly.

In combination with data in ASL-LEX, it will be possible to test hypotheses about a number of questions, including extending some that have already been examined using ASL-LEX alone. For example, Caselli & Pyers (2017) used ASL-LEX data to examine the iconicity and phonological neighborhood density of signing children's vocabulary development. With child-produced and child-directed frequency information to be made available in ASL Signbank, this kind of study can be extended.

As the ASL Signbank is used in annotating primary sign language data such as the SLAAASh corpus, it will make further research possible. Using multiple file searching functions of ELAN, it is possible to identify all instances in the corpus of signs of interest, which have been uniformly annotated because of the Signbank. As a further tool, we anticipate using lexical category information in Signbank to automatically tag SLAAASh data, which can be further tested in various ways. One only need consider the vast amount of research that has been made possible by the CHILDES database (https://childes.talkbank.org/) to anticipate the range of possible studies that will be forthcoming.

7. Conclusion

The actual usage of each sign in the corpora informs the ASL Signbank, from the most basic questions (which signs to include) to refinement of the postulated linguistic features. As our data set grows, our ability to answer these kinds of questions will improve. Lemmatization of ASL signs has never been done on a scale like this before - one that has been continually refreshed by actual usage data.

8. Acknowledgements

The research reported here was supported in part by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award number R01DC013578 and award number R01DC000183. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The work reported here is not possible without the entire SLAAASh research team, including Amelia Becker (who also provided feedback on drafts of this paper), Donovan Catt, Anna Lim Franck, Ardavan Guity, Carmelina Kennedy, Laura Mahan, Matthew Nardozza, Lettie Nazloo, Deborah Peterson, Lee Prunier, Doreen Simons, Phoebe Tay, Jacob Veeder, and all of the ASL SignBank actors. slla.lab.uconn.edu

The ASL Signbank was developed at Radboud University by Onno Crasborn, Wessel Stoop, Micha Hulsbosch, and Susan Even.

9. Bibliographical References

Battison, R. (1974). Phonological deletion in American Sign Language. *Sign Language Studies, 5*, 5-19.

Brentari, D. (1998). A prosodic model of sign language phonology. Cambridge, MA: MIT Press.

Caselli, N., Sevcikova, Z., Cohen-Goldberg, A., Emmorey, K. (2016). ASL-Lex: A Lexical Database for ASL. Behavior Research Methods. doi:10.3758/s13428-016-0742-0

Caselli, N. & J. Pyers. (2017). The road to language learning is not entirely iconic: Iconicity, neighborhood density, and frequency facilitate acquisition of sign language. Psychological Science. Vol 28, Issue 7, pp. 979 – 987.

Cassidy, S., Crasborn, O., Nieminen, H., Stoop, W., Hulsbosch, M., Even, S., Komen, E., Johnston, T. (2018). *Signbank: Software to Support Web Based Dictionaries of Sign Language*. Paper to be presented at the LREC 2018, Miyazaki, Japan.

Chen Pichler, D., Hochgesang, J.A., Lillo-Martin, D. (2015, March). BiBiBi Project ASL Annotation Conventions. Poster presented at "Digging into Signs Workshop: Developing Annotation Standards for Signed Language Corpora". University College London, London, United Kingdom (March 30-31, 2015).

Chen Pichler, D., J.A. Hochgesang, D. Lillo-Martin, & R. Quadros. (2010). Conventions for sign and speech transcription in child bimodal bilingual corpora. Languages, Interaction and Acquisition, 1(1), 11-40.

Chen Pichler, D., J.A. Hochgesang, D. Simons & D. Lillo-Martin. (2016). Community Input on Re-consenting for Data Sharing. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J.A. Hochgesang, J. Kristoffersen, & J. Mesch (Eds.), Workshop Proceedings: 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining / Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016) (29-34). Paris: European Language Resources Association (ELRA).

Crasborn, O. & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood & E. Thoutenhoofd (Eds.), Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the

Representation and Processing of Sign Languages. ELRA, Paris, pp. 39-43.
Crasborn, O., Bank, R., Zwitserlood, I., Kooij, E. van der, Schüller, A., Ormel, E., Nauta, E., Zuilen, M. van, Winsum, F. van., & Ros, J. (2016). *Linking Lexical and Corpus Data for Sign Languages: NGT Signbank and the Corpus NGT*. Paper presented at the The 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, Portorož, Slovenia
Crasborn O., R. Bank, I. Zwitserlood, E. Kooij, E. Ormel, J. Ros, A. Schüller, A. de Meijer, M. van Zuilen, Y. Nauta, F. van Winsum & M. Vonk (2017). NGT Signbank. Nijmegen: Radboud University, Centre for Language Studies. ISLRN: 976-021-358-388-6, DOI: 10.13140/RG.2.1.2839.1446.
Fenlon, J., Cormier, K, Schembri, A. (2015). Building BSL Signbank: The Lemma Dilemma Revisited. *International Journal of Lexicography*, 28(2), 169-206.
Johnston, T. (2001). The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics, 4*(1/2), 145-169.
Fenlon, J., K. Cormier, R. Rentelis, A. Schembri, K. Rowley, R. Adam, & B. Woll. (2014). BSL SignBank: A lexical database of British Sign Language (First Edition). London: Deafness, Cognition and Language Research Centre, University College London.
Fisher, J.N., J.A. Hochgesang, & M. Tamminga. (2016). Examining Variation in the Absence of a 'Main' ASL Corpus: The Case of the Philadelphia Signs Project. Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining. (bit.ly/PhiladelphiaSigns) 75-80. LREC, Portorož, Slovenia, May 28 2016.
Hochgesang, J.A. (2015, updated 2016) SLAAASh ID Glossing Principles and Annotation Conventions. Ms., Gallaudet University and Haskins Laboratories. http://bit.ly/2jGbPfU
Johnston, T. (2001). The lexical database of Auslan (Australian Sign Language). Sign Language & Linguistics, 4(1/2), 145-169.
Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. International Journal of Corpus Linguistics, 15(1), 106-131.
Lillo-Martin, Diane & Chen Pichler, Deborah (2008). Development of Sign Language Acquisition Corpora. In O. Crasborn, E. Efthimiou, T. Hanke, E.D. Thoutenhoofd, & I. Zwitserlood, (Eds.), Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora; 6th Language Resources and Evaluation Conference, 129-133.
Salonen, J., Takkinen, R., Puupponen, A., Nieminen, H., & Pippuri, O. (2016). Creating Corpora of Finland's Sign Languages. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Hochgesang, J. Kristoffersen, & J. Mesch (Eds.), Workshop Proceedings: 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining / Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016) (pp. 179-184). Paris: European Language Resources Association (ELRA).

10. Language Resource References

ASL Signbank (2018). Lexical database for American Sign Language. URL: http://aslSignbank.haskins.yale.edu.

Development of an "Integrative System for Korean Sign Language Resources"

Sung-Eun Hong[1], Seongok Won[1], Il Heo[1], Hyunhwa Lee[2]

[1]Korea National University of Welfare, Department of Sign Language Interpretation, Korea
[2]National Institute of Korean Language, Promotion Division of Special Languages, Korea
Sungeunhong2001@gmail.com, woonsok@knuw.ac.kr, heo1@knuw.ac.kr, lhh1127@korea.kr

Abstract

In 2015, the KSL Corpus Project started to create a linguistic corpus of the Korean Sign Language (KSL). The collected data contains about 90 hours of sign language videos. Almost 17 hours of this sign language data has been annotated in ELAN, a professional annotation tool developed by the Max-Planck-Institute of Psycholinguistics in the Netherlands. In the first phase of annotation the research project faced three major difficulties. First there was no lexicon or lexical database available that means the annotators had to list the used sign types and link them with video clips showing the sign type. Second, having numerous annotators it was a challenge to manage and distribute the hundreds of movies and ELAN files. Third it was very difficult to control the quality of the annotation. In order to solve these problems the "Integrative System for Korean Sign Language Resources" was developed. This system administrates the signed movies and annotations files and also keeps track of the lexical database. Since all annotation files are uploaded into the system, the system is also able to manipulate the ELAN files. For example, tags are overwritten in the annotation when the name of the type has changed.

Keywords: Korean Sign Language, corpus, annotation administration system, KSL resources

1. KSL Corpus

The KSL Corpus Project started to build the KSL Corpus 2015. It is the first effort to create a linguistic corpus of Korean Sign Language which fulfills the criteria of a modern corpus. This means that the corpus is machine readable and digital. The KSL Corpus Project collected sign language data from 60 deaf signers in the area of Seoul. The informants were invited in pairs and asked to complete 13 tasks which used different kinds of elicitation materials (cf. Hong et al, same volume). Each session with a pair of informants was three hours long that means the KSL Corpus Project has collected 90 hours of raw data and the project plans to collect more sign language data in other areas of Korea in the future.

2. Annotation of the Corpus Data

In the process of building the KSL corpus the KSL Corpus Project examined iLex – a database tool for integrating sign language corpus linguistics and sign language lexicography (Hanke & Storz 2008) as well as ELAN – a professional annotation tool developed by the Max-Planck-Institute of Psycholinguistics in the Netherlands. Although iLex offers many more advantages, the KSL Corpus Project decided to use ELAN because iLex would need much more IT knowledge in order to get things started and the KSL Corpus Project couldn't provide this kind of capacity at the beginning of the project.

The KSL Corpus Project recruited numerous hearing and deaf annotators. Unfortunately the KSL Corpus Project is not able to provide a location where the annotators could work together. That means all annotators do their work at home. This makes it hard to share and exchange thoughts and/or questions with each other during the annotation process. However, the annotators come together for the annotation training when they start the annotation and they come together once a week for an meeting where annotation problems are discussed and clarified. Based on this weekly annotation meetings the research project has documented the annotation conventions (National Institute of Korean Language, 2017). These conventions are the foundation of the KSL Corpus annotation and they help to keep the annotation process as consistent as possible.

Almost 17 hours of the collected 90 hours of KSL data have been annotated in ELAN[1] so far. The main goal of the first process of annotation was lemmatization – the classification or identification of related word forms under a single label. The KSL Corpus Project has followed Johnston (2008) by using ID glosses. The annotation can be seen as the first attempt in Korea to transcribe and annotate sign language data in a systematic way. Lemmatization is usually substantially easier when a reference dictionary or a lexical database exists (Johnston 2010). Since neither was available in Korea the annotator had to annotate and document the sign type at the same time. For each new found sign the annotator entered its type name on a google sheet and filmed him/herself signing the basic form of the annotated sign. The movie of the sign was stored on a cloud system and linked to the entry in the google sheet. This process resulted in a list of 2.400 different sign types.

The annotation environment in Korea is probably unique in several points. Not only do the annotators work separately from each other, but it is very difficult for the KSL Corpus Project to occupy annotators longer than 5 months since each phase of the project is only about 8 months long (March – December). Due to these circumstances the project is forced to employ numerous annotators for a short period of time. In the first phase of annotation the project had 23 annotators. Having so many annotators and having such an intense annotation phase we faced following problems. First, it was a challenge to manage the hundreds of movies and ELAN files distributing them among the annotators. Second, there was a strong need for a database instead of the sign types list

[1] The KSL Corpus Project has translated the interface of ELAN into Korean (Hangul). The Korean version of ELAN is available since version 4.9.3

in the google sheet. Third, if a name of a sign type was changed the annotators had to change the tokens in their ELAN files, but it was hard for the research project to control this step and to ensure consistency.

3. Integrative System for Korean Sign Language Resources (ISKSLR)

3.1 Introduction

Signbank is a lexical database for sign language resources which is used for sign language corpora such as the AUSLAN Corpus (Johnston 2001), the BSL Corpus (Cormier et al. 2012) and the NGT Corpus (Crasborn and Slöetjes 2014). Signbank was considered by the National Institute of Korean Language but was declined because Signbank is developed in Django web framework, which does not belong to the official recommended frameworks of the Korean government. Therefore the National Institute of Korean Language decided to develop an own system which fits to the needs and the setting of the KSL Corpus Project.

The ISKSLR is able to archive and administrate a) the KSL videos data with all its metadata b) the information about the informants, which were collected by a questionnaire before the data collection c) the elicitation material of each task and d) the ELAN annotation files.

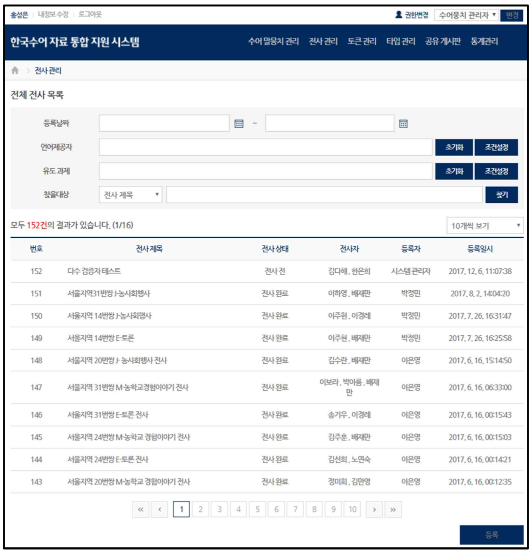


Figure 1: Searching page for annotation files

3.2 Annotation files

The ELAN annotation files can be searched by following factors: registration date of the annotation file, name or defining characteristic of the informant, sort of task, name of annotation file or name of the annotator (fig. 1).

If the annotation file is found, basic information is shown such as the name of the annotation file, ID number of the informants, linked KSL videos, all tiers within the annotation files and the information as to what extent each tier has been annotated. Figure 3 shows orange bars which represents this information (ISKSLR determines this by looking up the tag with the highest time code in the tier, e.g. if a KSL video is 10 min long and there is only one tag with the time code 00:05:00:00 it would falsely appear as if 50% of this tier would have been annotated). Furthermore one can see how many tags are in each tier and in what stage a tier would be. ISKSL distinguishes stages such as annotation in process, annotation completed, checkup in process and checkup completed. It is also possible to download the KSL videos in two different compression rates (low quality and high quality) as well as to download the ELAN file (fig 2).

Figure 2: Entry of an annotation file

If annotators are told to annotate he/she would download the KSL videos and corresponding ELAN file and start the annotation. After, but also during the annotation process the annotators upload their ELAN file to the ISKSLR. The ISKSLR accepts only annotation files and the corresponding tiers when these had been assigned to the annotator before by a project member. This inhibits the annotators from falsely deleting or editing existing annotations in other tiers. When the annotators upload their files to the ISKSLR, it is possible for the project member to view and check the annotations. The uploading also serves as a backup method.

3.3 Sign Type Database

Furthermore the ISKLSR keeps track of the sign type entries which the annotators created during the annotation. Currently a sign type entry contains the following information: gloss of the sign type, video showing the basic form of the sign type, sign type meaning, entry date, name of the annotator, who entered the sign type. There is no phonological information about the sign type. When an annotator finds a sign which is not listed in the ISKLSR yet, the annotator makes an entry in the sign type nomination list (fig. 3). The nominated sign types are either discussed in the annotation meeting or checked by a researcher. If a nominated sign type is accepted it appears in the ordinary sign type database. If a nominated sign type is not accepted the reason is noted the entry and the annotator can look it up. When a sign type is deleted or changes its name the ISKLSR is able to overwrite the corresponding tags in all annotation files within the system. The changes apply when the annotation files are uploaded and are visible when the files are downloaded again. Furthermore, every time an annotation file is uploaded the ISKLSR is able to find tags which do not match the sign type list. The annotator can edit the non-matching tokens within the ISKLSR without going back to ELAN. This control mechanism functions as a control for falsely annotated tags or spelling mistakes of the annotators.

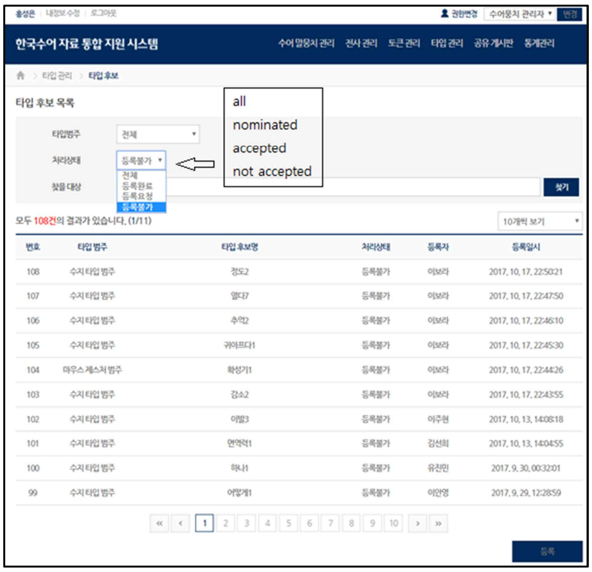


Figure 3: List of nominated sign types

3.4 Tags

It is also possible to search for the tags of a specific sign type. There are several different views one can choose to view a tag of a sign type. For example, one view shows only the corresponding video of the tag. Another view shows not only the corresponding video, but also the video of the opposite informant as well as the full shot video. Another view is able to present numerous videos at the same time (fig. 4), so the annotator can compare the video of the tags.

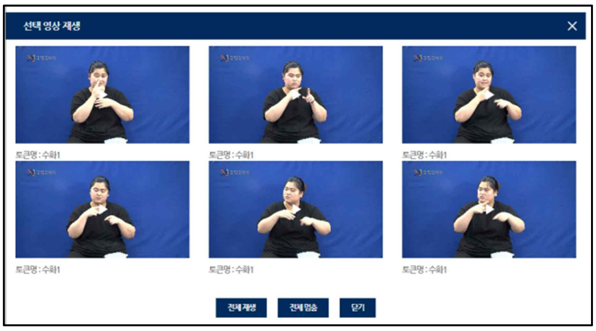


Figure 4: List of nominated sign types

3.5 Other functions

The ISKSLR also administrates things, which are usually handled in ELAN. The ISKSLR administrates tiers, linguistic types of tiers and controlled vocabularies. It is of importance to create tiers in the ISKSLR, not ELAN. This is because only then it is possible for the ISKSLR to manipulate the tags in the ELAN files.

The ISKSLR offers simple statistics, too. For example, it is possible to get an overview of all annotation files and to see how much of the files have been annotated, how much are in process and so on. Also an overview of mismatched tags, work load of the annotators, among other statistics are available (fig. 5).

Figure 5: Simple statistic functions

The ISKSLR also stores documents like annotation conventions and annotation minutes and has space for general announcements as well as for questions of the annotators. These rather simple functions are essential to the annotators since they see each other only once a week.

Recently the developers of the ISKSLR have edited ELAN in such way that it is possible to view the sign type list of the ISKSLR within ELAN. After ELAN is opened the annotator is asked to login in the ISKSLR. The annotator can now either type the name of the sign type (like in the past) or to choose from the ISKSLR sign type list, which also presents the sign type video (fig. 6).

Figure 6: Sign type list of the ISKSLR appears in ELAN

4. Conclusion

The ISKSLR was developed to meet the special needs of the KSL Corpus Project. One characteristic of this research project is the fact that all annotators are working at home and there is a high fluctuation. The ISKSLR is trying to compensate for this by providing a structure to store, administrate and assign annotation files effectively. A second function of the ISKSLR is to improve the consistency of the annotations. All annotation files have to be uploaded and changes in the sign type list will apply directly to the ELAN files. There is also a mismatch function which finds all tags which cannot be assigned to the sign type entries. To bring more transparency, the ISKSLR labels what stage an annotation file or even tier is in at that moment (annotation in process, annotation completed, checkup in process and checkup completed), but since the ISKSLR is not able to label the annotation files automatically it turns out to be an additional task to the annotators. Although small annotation changes such as the re-assigning of a tag, or change of a sign type name can be done in the ISKSLR, mostly it is necessary to go back to the ELAN file and make the changes directly (this is especially true for segmentation changes). Each time the annotators have to download the associated annotation file and link them with the sign language videos. This process is simple when an annotator is assigned only one annotation file, but gets very time consuming when you are dealing with numerous files.

It is planned to add some more functions in the ISKSLR. The statistic functions are expected to get more complex and it is hoped that the simple communication functions in ISKLSR get more deaf adequate such as an integrated video chat function between the annotators.

5. Acknowledgements

This work is supported by the National Institute of Korean Language. The ISKSLR is developed by the software company DIQUEST, Korea.

6. Bibliographical References

Cormier, K., Fenlon, J., Johnston, T., Rentelis, R., Schembri, A., Rowley, K., Adam, R., and Woll, B. (2012). From corpus to lexical database to online dictionary: Issues in annotation of the BSL Corpus and the development of BSL SignBank. *5th Workshop on the Representation of Sign Languages: Interactions between Corpus and Lexicon [workshop part of 8th International Conference on Language Resources and Evaluation*, Turkey, Istanbul LREC 2012. Paris: ELRA. pp. 7–12.

Crasborn, O., and Slöetjes, H. (2014). Improving the Exploitation of Linguistic Annotations in ELAN. In N. Calzolari, K. Choukri & et al. (Eds.), *Proceedings of LREC 2014, Ninth International Conference on Language Resources and Evaluation*. Paris: ELRA.

Hanke, T. Storz, J. (2008). iLex: A database tool for integrating sign language corpus linguistics and sign language lexicography. In Proceedings of the 6th Language Resources and Evaluation Conference (LREC) Marrakech, Morocco, pp. 64-76.

Hong, S.-E., Lee, H., Lee, M.-H., Byun, S.-I. (In Press). The Korean Sign Language Act. In Maartje De Meulder, Joseph J. Murray and Rachel McKee, editors, Recogninzing Sign Languages: An International Overview of National Campaigns for Sign Language Legislation and their Outcomes. Multilingual Matters.

Johnston, T. (2001). The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics* 4.1-2 (2001): 145-169.

Johnston, T. (2008). Corpus linguistics and signed languages: No lemmata, no corpus. In O. Crasborn, T.Hanke, E. D. Thoutenhoofd, I. Zwitserlood, & E. Efthimiou (Eds.), Construction and exploitation of sign language corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages (LREC) (pp. 82–87). Paris: ELRA.

Johnston, Trevor. 2010. From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics* 15(1). 106–131.

Jung, Heewon (2016). Korean Sign Language Policies: Trends & Outlook. International Conference "Sign Language and Social Communications", National Institute of Korean Language and Korea Deaf Association, Seoul.153-167.

National Institute of Korean Language (2017). 한국수어 말뭉치 연구 및 구축. Research and Building of the KSL Corpus.서울. 국립국어원. Seoul, National Institute of Korean Language.

Raising Awareness for a Korean Sign Language Corpus among the Deaf Community

Sung-Eun Hong[1], Seongok Won[1], Il Heo[1], Hyunhwa Lee[2]
[1]Korea National University of Welfare, Department Sign Language Interpreting, Korea
[2]National Institute of Korean Language, Promotion Division of Special Languages, Korea
Sungeunhong2001@gmail.com, woonsok@knuw.ac.kr, heo1@knuw.ac.kr, lhh1127@korea.kr

Abstract

This paper contains strategies that need to be implemented before the sign language community can be involved in corpus work to raise awareness for the need of corpus work. The Korean Sign Language (KSL) Corpus Project began in order to create a linguistic corpus with 60 deaf native and near-native signers from the area of Seoul. In the process of building the KSL Corpus by collecting sign language data and annotating it the project was faced with the challenge that the concept of corpus was completely new to the Korean deaf community. The KSL Corpus Project developed three strategies in order to inform and explain what the KSL Corpus is about. First, the research project produced numerous KSL videos and posted them on social networking websites in a weekly rhythm. Second, the project organized a workshop, where only deaf people were invited to participate. Third, the KSL Corpus project selected prominent deaf people who were schooled and provided with corpus materials in order to inform others about KSL Corpus by connecting to their friends and families. The experiences and outcomes of the above strategies are of special importance since the data collection of the KSL Corpus is still in process.

Keywords: Korean Sign Language, corpus linguistics, involvement of deaf community, deaf workshop

1. KSL Corpus

The Korean Sign Language (KSL) is the sign language of the deaf people in South Korea. Due to the Korean Sign Language Act, which was enacted in 2016, KSL has now gained legal recognition and is the official language of deaf people in South Korea with its own rights and an equal status to the spoken Korean language (Hong et al., in press). One of the aims of the KSL Act is to protect the linguistic rights of deaf communities and KSL users through a variety of KSL-related research projects (Jung, 2016). One of these projects is the KSL Corpus Project which is funded by the National Institute of Korean Language and carried out by researchers from the Korea National University of Welfare.

The KSL Project has collected sign language data from 60 deaf signers in the area of Seoul. The spontaneous and elicited data (90 hours of raw data) has been tagged and annotated using ELAN (Wittenburg et al., 2006). ELAN is a professional tool used to create complex annotations on video and audio resources. It was developed by the Max-Planck-Institute. So far only the sign language data of informants from Seoul have been collected. It is the intention of the project to collect more KSL data of other areas in the future.

1.1 Data Collection

1.1.1 Informants

60 deaf native and near-native signers from the area of Seoul were invited in pairs. The informants have been recruited with the help of the Korea Association of the Deaf and its 25 offices in Seoul. In the process of selection we realized how important it was to recruit the informants with the help of the Deaf Association. The Korea Association of the Deaf has an important and central role within the deaf community and was able to reach out to the deaf people. In addition, the informants participated in the research project just because it was supported by the Deaf Association.

All informants were at least 19 years old. Their most used and most comfortable language is KSL. They have graduated from a deaf school (exceptions were made, when the informants are a child of deaf parents). They have either lived in Seoul for at least 10 years or they have lived close to Seoul and work in Seoul for at least 10 years. Lastly, they meet deaf people at least three times a week. Informants have been prioritized in the process of selection when they had deaf parents, siblings or partners and if they had acquired KSL before they entered school.

1.1.2 Elicitation Materials

The elicitation materials used in the KSL Corpus contain pictures, photographs, movie clips, animations, topics for an open conversation, signed videos and a combination of pictures and written words. These types of stimuli are processed in 13 tasks which are mainly based on the elicitation material of the DGS Corpus (Nishio et al., 2010). Most of the tasks from the DGS Corpus Project have been adapted to the Korean deaf culture. The elicitation materials have been tested twice before the data collection took place. Each task is introduced and explained in sign language. In order to ensure that all informants receive the same input, the instructions were presented in a video.

1.2 Annotation and Translation

Each session of this naturalistic, controlled and elicited signed language sample has a length of about three hours. This means that the complete recordings contain about 90 hours of sign language data. All the video material has been cut into the length of each task, converted, compressed into MPEG format and synchronized in order to use the data with ELAN. About two thirds of the data was translated into Korean by competent KSL-interpreters. And almost 17 hours of the KSL data has been annotated in ELAN. The annotation of the KSL Corpus Project can be seen as the first attempt in Korea to transcribe and annotate KSL data in a systematic and scientific way. The

KSL Corpus Project recruited numerous hearing and deaf transcribers. But perhaps different than sign language annotators in countries like Germany for example, the Korean annotators work at home and participate in a weekly annotation meeting where annotation problems are discussed and clarified.

2. Raising Awareness of the Need for a KSL Corpus among Deaf Community

When the data collection of the KSL Corpus Project started in 2015, we assumed that corpus work was unknown to deaf people just like the majority of hearing people. We explained to the informants what a corpus is and why a KSL Corpus is of great importance to the sign language community. However, we had the desire to reach more people than just the informants. Our goal was to raise awareness for a KSL Corpus and to let the deaf community know that the KSL Corpus is not just a research object of (hearing) linguists. But it is a language resource which belongs to the deaf community and deaf people should be proud of it, because KSL is unique and an independent full-fledged language with its own structure and grammar. The building of the KSL Corpus comes together with the KSL Act which was enacted in 2016 (Hong et al., in press). Because of this legislation, the KSL Corpus has received much attention from the deaf community. The strategies developed by the KSL project contain three different approaches, which all have the aim to inform and draw interest to the KSL Corpus.

2.1 KSL Videos

The KSL Project produced 16 KSL videos in order to reach out to the deaf community. First of all, a deaf member and a CODA member of our research team created six short KSL videos in which both were shown in a conversation about the following issues: What is a KSL corpus? What is annotation about? What does a data-collection session look like? Why is a corpus important? What are the advantages of having a KSL corpus? How can a corpus be used? Why is a KSL corpus an important contribution to deaf Culture? Why is it important that deaf people get involved in corpus research? The video clips each 5-6 min long were not recorded in a studio but were recorded in a coffee shop to keep a relaxed atmosphere.

A: CODA researcher
B: Deaf researcher

A: Hi!
B: Hi!
A: What are we doing here?
B: We are sitting in front of the camera because there is a research project from the NIKL and the Korean University of Welfare. It' called KSL Corpus Project and we are here to talk about it und this video will be posted on Facebook. We hope that our video can inform about the project and communicate with people who are curious. It is possible to post questions under the video and we will answer the questions.
A: Oh, I see! Well, there is not much I know…what could we talk about? (…) How did you sign corpus?
B: I signed it this way. There is no KSL sign for corpus yet, but I use this sign. In the Netherlands they sign it like this and in Germany they sign it like this. I will use this sign.
A: Ok (continues to repeat the sign for corpus). I understand. So, what is a corpus? I have heard about it recently, but to be honest I don't have a clue what it means.
B: Corpus is a collection of a huge bunch of language data. But you cannot just go and collect it the way you want. There have to be certain criteria and you also have to transcribe them.
A: Stop! What you mean by transcribe?
B: To transcribe means in an easy way to write things down somehow. You could write down a sign by using a pen and sheet of paper, but if you do it that way, it's not really searchable. So the writing should be done on a computer.
A: Wait, lots of my friends have recorded KSL data with their cameras. Do they all have a corpus then?
B: No. Having KSL videos doesn't mean you have a corpus. A Corpus means…

Figure 1 : Part of the English translation of a KSL video

The deaf colleague explained the above points in KSL and the CODA colleague asked questions (also in KSL) and stopped the deaf colleagues when things got too complicated. The language of both was kept easy and casual and technical terms were avoided as much as possible (see fig. 1).
These videos were uploaded on our Facebook page in a weekly rhythm (see fig. 2). Facebook is the most used social media within the deaf community in Korea and the numbers of people who saw, commented and shared the KSL videos were very satisfying to us. The viewers were asked to formulate questions about the content and the research team answered them on Facebook openly. The research project posted numerous other KSL videos in order to announce events (e.g., deaf workshop), give updated information about the research project or to find deaf informants or deaf annotators for example.

Figure 2: KSL video posted on the project Facebook page

2.2 Deaf Workshop

The second strategy was to organize a workshop in which only deaf people were invited to attend (fig. 3). This workshop contained three parts: A KSL talk on the importance of a KSL corpus, small group work on different aspects of a KSL corpus and a poster session. The KSL talk explained the importance of a KSL corpus in detail. It was possible to explain things in context and show examples by presenting visual material which aided in the understanding of the concepts being presented. While the KSL videos were kept very casual, the talk on the deaf workshop had an academic character. After the talk, the workshop participants were asked to join small groups (fig. 3) which discussed issues such as: Do we need standardization in KSL? What is the role of the KSL corpus concerning standardization? Does KSL corpus succeed in language documentation as well as documenting the culture of the deaf? What happens if "wrong" signs are included in the KSL corpus? These topics were chosen because we had the feeling that these things were currently discussed in the deaf community which could be also seen by the comments on Facebook. Especially standardization is always an issue among deaf people in Korea, because various sign language dictionaries in the past claimed to present standardized signs, but couldn't support it with language data (Lee, 2017).

Figure 3: Official poster of the Deaf Workshop

All small groups (fig. 4) were moderated by a deaf or CODA project member, who received special training. The small groups were equipped with marker pens and white posters and were asked to draw the results of their discussion. This method seemed more deaf-friendly than using written language. After the discussion time a participant of each small group presented the results by showing and commenting on the small group poster to all workshop participants.


Figure 4: Small group session at the Deaf Workshop

The breaks in the workshop were used as a poster session. The research group prepared several posters with different topics. The issues were presented simply and visually. Additionally, deaf project members explained the posters.

Figure 5 shows one of the workshop posters, which explains in what ways a KSL corpus benefits the deaf community. Issues like regional or age variation of signs, corpus-based sign language research, preservation of KSL, corpus as a language resource, and other relevant concerns are presented visually on the poster. Figure 6 shows the studio setup of the data collection.

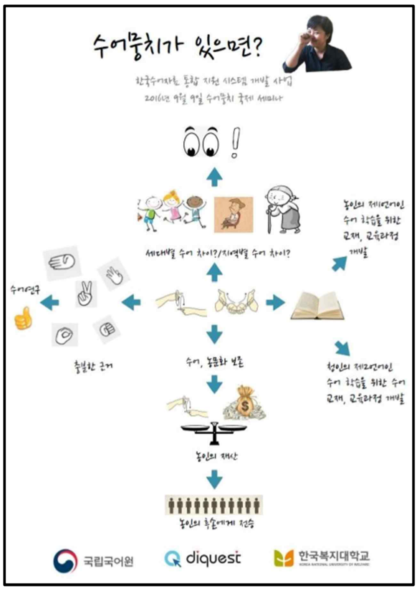


Figure 5 : Poster at the Deaf Workshop

The deaf workshop received much attention and positive feedback in the deaf community[1]. The decision to exclude hearing people (which was taken seriously, the hearing researchers were also not allowed to participate actively at the workshop) helped many deaf people to express themselves curiously, critically and openly.

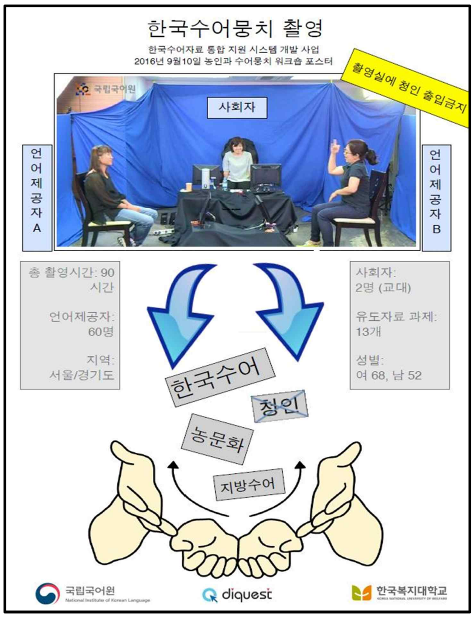


Figure 6 : Poster at the Deaf Workshop

[1] Deaf participants were asked if they liked and what they liked about the workshop. Their answers were filmed and edited to a video which was posted on Facebook.

2.3 Schooling of Prominent Deaf People

The third strategy had the aim to educate prominent deaf people and to send them to their deaf communities to inform other deaf people about sign language corpus work. We chose deaf people who had leadership roles and/or a significant influence in the deaf community. These chosen prominent deaf members participated at the deaf workshop and were additionally educated by the research team. We provided them with teaching material such as KSL videos, posters, pictures and PowerPoint presentations and asked them to go to their circle of friends and acquaintances and to meet 3-5 deaf people for about an hour to inform them about what a sign language corpus is about and what it can accomplish for the deaf community. Each educated deaf member was asked to arrange several of these sessions in places of their choice (e.g., at home, at church, in coffee shops) and to record one minute of each session on video (see fig. 7). This strategy was developed for two reasons: first, this way of information also reaches out to deaf people in rural areas so long as the prominent deaf person comes from a rural area. Since our deaf workshop took place only in Seoul, there was no possibility for us to reach deaf people in the provinces. Second, this method reached people who weren't always well connected to the mainstream deaf community. Through this method, people got to know about the KSL corpus in a very private and intimate atmosphere from a person they trusted.

Figure 7: Deaf person telling his friends about the KSL Corpus

3. Conclusion

All three strategies had the aim to raise awareness of a sign language corpus in the deaf community. The KSL video method reached mostly young people because the videos were spread through an internet medium. This also had the advantage of being locally unbounded. The deaf workshop was a great opportunity to assemble numerous deaf people in order to discuss corpus-related matters. It also turned out to have a great impact on the deaf community because it was the first event in the framework of sign language research excluding hearing people and showed that a sign language corpus is about deaf people (and not about hearing people doing sign language research). Schooling prominent deaf people was an approach to meet more specific deaf communities in a private surrounding. Depending on the person who was educated, the groups of people he or she would meet would be specific to their community. An elder deaf person would probably meet elder people and a deaf person from the country would be meeting people from the country. The feedback of the deaf participants concerning the above methods was highly positive. Many deaf people were thrilled to experience sign language research in a visual and deaf-friendly way. The main point of the introduced strategies is that all these things cannot be done without very experienced deaf researchers. Although the concept of the strategies was developed together in the project, the three strategies couldn't have been carried out without the professional role of the deaf researchers. It is not a new finding that promoting and training of deaf researchers is beneficial. But this is especially true when the aim is to involve the deaf community in corpus work. The KSL Corpus Project aims to train more deaf persons and to find the deaf leaders who can actively involve the deaf community to the KSL Corpus.

4. Acknowledgements

This work is supported by the National Institute of Korean Language. We acknowledge our deaf colleagues Eun-Young Lee and Kang-Suk Byun, whose role in this part of research was essential.

5. Bibliographical References

Hong, S.-E., Lee, H., Lee, M.-H., Byun, S.-I. (In Press). The Korean Sign Language Act. In Maartje De Meulder, Joseph J. Murray and Rachel McKee, editors, *Recogninzing Sign Languages: An International Overview of National Campaigns for Sign Language Legislation and their Outcomes*. Multilingual Matters.

Jung, H. (2016). Korean Sign Language Policies: Trends & Outlook. International Conference "Sign Language and Social Communications", National Institute of Korean Language and Korea Deaf Association, Seoul.153-167.

Lee, H. 이현화 (2017). 한국수어 정비 사업: 한국수어사전(한국수어의 기록). Korean Sign Language Inspection Project: Korean Sign Language Dictionary(Documentation of Korean Sign Language). 새국어생활, *New language life*. 27-2, 국립국어원, National Institute of Korean Language, pp. 67-84.

Nishio, R., Hong, S.-E., König, S., Konrad, R., Langer, G., Hanke, T., Rathmann, C. (2010). Elicitation methods in the DGS (German Sign Language) corpus project. In *Proceedings of the 4th workshop on representation and processing of sign languages: corpora and sign language technologies*. European Language Resources Association, pp. 178-185.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pp. 1556–1559.

Publishing DGS corpus data: Different Formats for Different Needs

Elena Jahn, Reiner Konrad, Gabriele Langer, Sven Wagner, Thomas Hanke
Institute of German Sign Language and Communication of the Deaf, University of Hamburg
Gorch-Fock-Wall 7, 20354 Hamburg, Germany
E-mail: {elena.jahn, reiner.konrad, gabriele.langer, sven.wagner, thomas.hanke}@uni-hamburg.de

Abstract

In 2010-2012, the DGS-Korpus project collected a large corpus of German Sign Language (DGS). Now, a substantial subset of the data is published, namely the Public DGS Corpus. We describe the considerations and decisions taken regarding what part of the data is to be made public, the necessary quality assurance measures to the data preparation as well as the formats of the published data. The corpus is published in three different ways in order to fulfil the needs of a variety of different users. First of all, the data is made available to the language community whose members allowed us to share their recorded language. In addition, we hope that a large number of non-scientific users with various backgrounds will find the data useful. Last but not least, we aim to make the data attractive for users with a scientific background and provide the possibility to conduct studies based on it, irrespective of whether they are familiar with DGS or not.

Keywords: DGS (German Sign Language), corpus building, involvement of the language community, long-term accessibility of sign language data

1. Introduction

In the last ten years, the number of large-scale sign language corpus projects has been growing, in line with the understanding that corpora should form the underpinnings for many research areas, one of them being lexicography. At the same time, there is an increased awareness that the respective language community should benefit from the collected data. On the other hand, funding of corpus work has often been related to specific research questions and not the general usefulness of the data. Large, representtative sign language corpora only recently started to emerge. Thus, it remains a key issue for any sign language corpus work to make the data accessible. Ideally, published corpus data are used frequently by different kinds of users. Therefore, a low-threshold access to the data, suitable for non-scientific users as well as users with a scientific background, should be a requirement.

For sign language corpora, publishing and making data publicly accessible is extremely challenging for ethical reasons (complete anonymization of the informant is not possible in video data), technical reasons (storing video data and keeping them technically up-to-date), historical reasons (the lack of standardised procedures well accepted in the community) and due to a matter of resources (personal and financial) and sustainability.

As with any minority language research, the success of a sign language corpus project strongly depends on the participation and involvement of the language community. Not only are members of the Deaf Community needed in order to gain samples of natural signing by a native signer, but also is the expertise of native signers needed in the process of reviewing translations and annotations. Corpus-based research on sign languages is thus impossible without the help of the Deaf Community. In acknowledgement of the Deaf Community's contribution they should be given continuing access to the data even beyond the period of data collection and processing.

In addition, it is also important that linguistically motivated research on sign languages is facilitated by means of providing corpus data that is suited for publication. However, the detailed exploration of a sign language on basis of a sufficiently large corpus hinges on technical requirements. Therefore, corpus based research on sign languages is a relatively young area of research where the scientific community is still striving for standards. For the aforementioned reasons, it cannot be taken for granted that sign language corpora are published at all. However, technical advancements nowadays facilitate the storage and publication of data online and thus enable projects to share not only their results but also their data. This open-access policy has some major advantages: "When data is accessible to other researchers, research outcomes can be checked by colleagues working in the same field; cross-linguistic studies are facilitated because similar data sets can be recorded for additional languages; the creation of new research groups and the work performed by a single researcher (as for dissertation projects) will become easier because part of the data collection effort can be skipped; finally, seeing in which way other data sets have been collected can lead to the gradual improvement in methodologies for the whole field." (Crasborn et al. 2007: 542)

The DGS-Korpus project is a long-term project of the German Academy of Sciences with two goals: building a reference corpus of DGS and compiling a corpus-based dictionary DGS – German. The raw video data, metadata, and annotations are stored in the iLex database (hereafter iLex), an annotation tool and lexical database that was designed as a multi-user application for annotation and lemmatisation of sign language data (Hanke 2002, Hanke/Storz 2008). Basic annotation includes a translation into German, lemmatisation, and annotation of mouthings/mouth gestures. Detailed annotation is concerned with differentiating between morpho-syntactic inflection, modification, and phonological variation as a basis for the lexicographic analysis and description of signs. Data can be retrieved with customised lists, filters, and queries using SQL. Furthermore, map functions and graphs are integrated, so that e.g. regional distribution of sign variants or variation between age groups can be visually displayed (see Hanke et al. 2017, Langer et al. 2018). Upon request, acces to the corpus data in iLex (as well as the software) is available to researchers outside the project.

Out of the 560 hours of DGS collected in 2010-2012 (Nishio et al. 2010), a subset of about 50 hours is made available as the Public DGS Corpus. It contains almost 400 episodes covering 18 different elicitation tasks ranging from experience reports of Deaf individuals to discussions, story retellings and jokes (see section 2).

We assume that different user groups will address the published data with different expectations. Deaf individuals may be interested in seeing their grandparent generation talk about earlier times, hearing learners of DGS may want to see signs in context or different styles of signing, sign language instructors may search for course material, interpreters for regional variants or DGS equivalents to technical terms, and linguists for appropriate natural DGS signing to conduct crosslinguistic research. Users should be given the possibility to utilise the data as a valuable basis for the investigation of many different questions concerning both the language itself and the language community. To address the different needs and interests, the data is made available via three formats: meine-dgs.de (see 3.1), the Research Portal (see 3.2), and ANNIS (see 3.3).

Providing different formats to access the Public DGS Corpus hopefully contributes to inviting many people to utilise the data. This is supported by encouraging the interaction between users and providing the possibility to report annotation mistakes to the DGS-Korpus team (see 4.).

2. Public Corpus Content

The Public DGS Corpus contains about 50 hours of signed conversations of pairs of interlocutors. The videos are presented bipartitely, with the interlocutors side by side. (In the studio setup, interlocutors were placed facing each other. For more information see Hanke et al. 2010.)

2.1 Prioritising and Selection of Video Material

The videos were carefully selected in order to

- be balanced for region, sex, and age,
- include all elicitation tasks (with the exception of the task "Sign names" for anonymisation reasons),
- cover a great variety of topics,
- cover different styles of signing,
- include each informant at least once.

The corpus shows 327 out of 330 informants (only three informants did not approve the online publication of their data).

The selection process started with a rating of elicitation tasks, in which each project team member rated each task with respect to its importance for the deaf community. As a result, eight tasks were prioritised (in descending order): "Experiences as a Deaf person", "Joke", "Free conversation", "Discussion", "Subject areas", "Experience reports", "Region of origin" and "Deaf events".

With the exception of the task "Joke" (that is, compared to other tasks, rather short), these tasks were proportionally allocated to the planned 50 hours of the public corpus. The remaining tasks were included only exemplarily. We excluded the task "Isolated items" which has a strong lexicographic interest (variation) from the public corpus. In sum, over 47 hours of the videos are selected from the seven tasks listed above, 1.7 hours from remaining tasks, and 2.4 hours from "Jokes".

Within some tasks we presented several stimuli to the informants, e.g. in the task "Subject areas" informants were given four different subjects from which they had to chose two for discussion. These parts (hereafter subtasks) were treated as independent units for annotation workflow. In the next step, subtasks from different informants were selected. For each subtask we revised, among other things, whether the content was appropriate for publication, the style of signing was comprehensible, the video was pleasant to look at, or whether technical difficulties occured during post-production.

2.2 Processing Steps

2.2.1 Indexing Content for Thematic Access

In order to facilitate a thematic access to the videos each selected subtask from the prioritised task list (see above) was indexed for content. A subtask could have one or several descriptors assigned to, but the majority of subtasks were indexed for several descriptors. The descriptors constituted a controlled vocabulary list of about 530 items. Each of these descriptors was assigned to one (or several) of 35 topics. These topics are an extended version of the originally 26 subject areas that were targeted in an elicitation task specifically designed to cover the basic vocabulary of DGS. On the website meine-dgs.de the videos can be filtered by chosing one of the topics (button "Alle Themen"), the more specific descriptors are then displayed below the video screen to facilite a more precise selection according to interest of the user. In the Research Portal the topics are listed under the column "Topics".

2.2.2 Translation into English Version

With the exception of the task "Joke" all selected subtasks were translated into German and lemmatised as part of the basic annotation. In a second step, they were translated from German into English. This enables researchers knowing neither DGS nor German to browse the content of the public corpus (in the research portal and in ANNIS). In addition, the German glosses were also translated into English and are displayed in the English version of the online transcript view.

2.2.3 Blackening and Anonymisation

We spent some effort to anonymise parts of the signing that should be exempt from the online publication. If stretches to be anonymised were too long, the subtask was not selected for inclusion in the Public DGS Corpus. In other cases, we decided to shorten the subtask, mostly at the beginning or the end. Finally, we have some cases left where stretches had to be blackened within a subtask (in general only a few seconds). In order to anonymise personal data of the informants or third persons (names, dates like birthday, or geolocations) we tagged these sequences, decided whether hands, mouth, or both had to be blackened and generated rectangle coordinates as annotations. These coordinates had to be checked manually in the frontal and profile view of the informants. When exporting the movie files the designated blocks were rendered black. Besides the videos, also translation texts, mouthing annotations and glosses hat do be identified and processed in order to produce anonymised texts and annotations (for details see Bleicken et al. 2016).

2.2.4 Editorial Steps

The publication of a sign language corpus requires additional steps not crucially necessary to work with the data in-house.

A built-in spell checker in iLex (for German and English) supported the annotators when aligning the German translations. Glosses and mouthings were checked manually.

Further on, we checked translations against lemmatisation and mouthings in order to reach a high consistency of the

annotations. This checking helped to fill translation gaps and revise unclear passages or to correct token-type mismatches and mouthings. The experience of the Deaf team members, Deaf students, and CODAs was indispensable and most valuable in this step.
Also, inconsistencies in the segmentation of subtasks, translations, tokens and mouthings/mouth gestures had to be checked, e.g. a translation tag should not start before or end after a subtask tag, a translation tag should not start or end in between a token or mouthing/mouth gesture tag. Overlapping translation tags of informant A and B with no significant signing had to be corrected.
Last but not least, each processing step helps to improve the quality of the annotations. Annotators comment and give feedback to translation and lemmatisation that were reviewed. This checking procedure has the drawback that tags were changed or comments were added after the corresponding annotation or checking step was already done. But this seems to be unavoidable when working with a team of 15 colleagues and over 30 student co-workers.

2.2.5 Persistent Identifier

We provide persistent identifiers for individual transcripts to make them quotable in a revision-savvy way.

3. Different Formats for Different Needs

The formats in which the Public DGS Corpus is distributed are the following:

- The website meine-dgs.de is a low-threshold access to the data. In this portal, videos are presented together with German translations as subtitles. Here, the focus is on content-related access.
- The Research Portal provides the video data with basic annotations as well as metadata on the informants for linguistic and related research. Annotation data (in German and English) is made available for download in ELAN and iLex format, or can be previewed in the web browser.
- In order to also provide easy access for researchers not familiar with annotation environments prevalent in sign language research, but with corpus tools in general, we also plan to make our data accessible via ANNIS (ANNotation of Information Structure; http://corpus-tools.org/annis/; Krause & Zeldes, 2016). ANNIS is a corpus query tool for visualization and querying multi-layer corpus data that comes along with its own query language.

3.1 meine-dgs.de

The first publication format, meine-dgs.de, is a website where users can watch the signed conversations or narratives with subtitles showing the translations into German, except jokes. In addition to the main page with the videos, the website contains information about the project, license terms and a page where the videos can be filtered for region, age groups, dialogue formats and main topics.
The website meine-dgs.de is meant to address users that are interested in the content of the conversations and narratives. It provides a low-threshold access to the data and is thus suitable for both users without a scientific background and users with a scientific background that would like to get familiar with the data. Also, users with a scientific background that is not linguistics or sign languages might find the data interesting, e.g. for studies concerning Deaf Culture or the way in which Deaf individuals have experienced decisive events. DGS is known to have regional variants, therefore users might want to search for videos from specific regions only.
The appearance of the website is as follows. On the main page, users can decide for jokes only ("Sammlung Witze" leading on a page with the format "Witze" (88 jokes) preselected), for all subtasks ("Sammlung Gespräche" with no format preselected), and the possibility to preselect the region via a map ("Sammlung Regionen"). For each video a short description is provided which contains information about the region (city or geographical area) where the conversation has been filmed (and the interlocutors are rooted), the dialogue format and the topics. This information is meant to help the user to get an overview over the data and select the most interesting videos.
The video contains subtitles that can be turned off and on at will. Below the video on the left, a mistake button ("FEHLER?") is implemented that allows for a non-public indication of mistakes to the DGS-Korpus team. On the right side a share button ("VIDEO TEILEN") enables to share the respective video in various social networks and platforms.

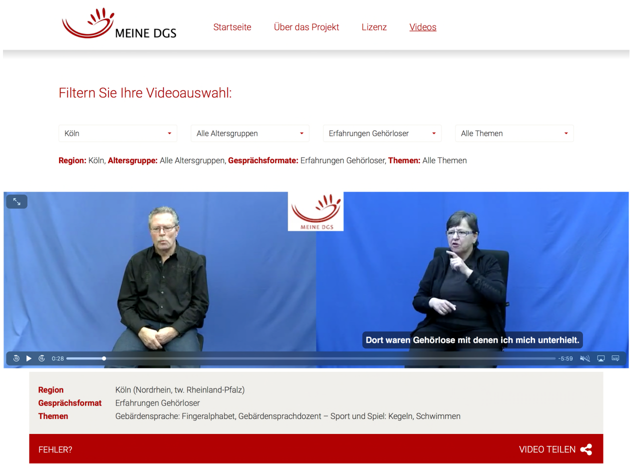


Figure 1 : meine-dgs.de

The general aim of the publication of the data on the website meine-dgs.de was to allow the user to concentrate on the content of the signed conversations or narratives. With this format we place importance on a low-threshold entry point for any interested person. Also, we hope to provide valuable data for learners and teachers of DGS who might use the videos for practice purposes. Researchers that are interested in getting an overview of the content of the conversations and the recording situation might find the site helpful, too. Also, the website serves as an open archive for language, culture, and history of Deaf individuals.

3.2 Research Portal

This portal is made for users with a scientific background who are interested in the content of the conversations and narratives, but with a focus on the language DGS itself.

Like meine-dgs.de, the Research Portal is accessible without prior registration. As it is supposed to address an international audience, the website is in English. It provides the same videos, here without subtitles, but augmented by annotations. It starts with a list of "Transcripts" (i.e. the subtasks), offers a "Types" list with all types used for lemmatising the tokens in the public corpus, links to the "Annotation Conventions" and informs about the conditions of use ("License"; see 5.). In the header a banner displays all informants.

The body shows a list of all subtasks. Instead of filters the subtasks are listed by the transcript name like "dgskorpus_ber_01" coding the region (ber=Berlin) and a running number for the elicitation session. Further codes are: fra (Frankfurt), goe (Göttingen), hb (Bremen), hh (Hamburg), koe (Köln), lei (Leipzig), mst (Münster), mue (München), mvp (Mecklenburg-Vorpommern), nue (Nürnberg), sh (Schleswig-Holstein), and stu (Stuttgart). Thus, the filter "Region" is dispensable. Age group, format, and topics are further columns in this list. The next columns contain icons to download annotation and video files. Annotation files are offered for iLex and ELAN import and are more extensive than the online transcript as they include both German and English translations and glosses, and additionally HamNoSys notations of the citation form of the types. Video files (h.264 codec, 640x360, 50 fps) are provided not only for informant A and B, but also for a total perspective with both informants in profile view and the moderator in the middle.

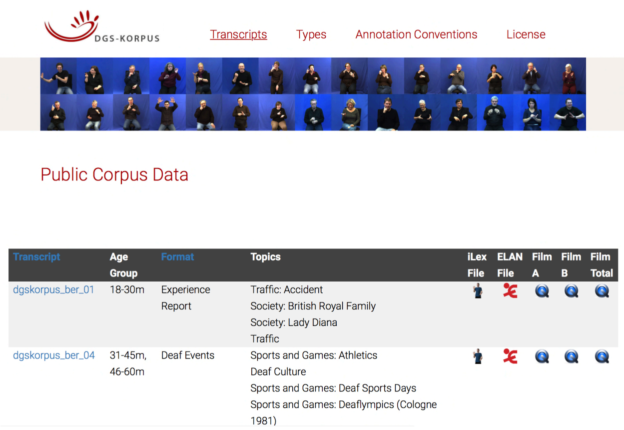


Figure 2 : Research Portal main page

By clicking on the "Transcript" name, videos and annotations can be browsed by an online transcript view (with the possibility to switch between a German and an English version). In this way, it differs from other access formats to sign language corpora. The online transcripts may be of interest also for users without a scientific background. It gives everyone a glimpse on how basic research in sign language corpus linguistics looks like and makes the results of our work transparent.

In the online transcript view, the videos with both informants are displayed at the top, with the transcript beneath. The annotation tiers are arranged in a vertical grid with a top-down timeline (as opposed to a horizontal grid many researchers may be used to). The timeline shows timecode start and end for each tag.[1] Three annotation tiers for each informant exist: Translation, Lexeme/Sign, and Mouthing/Mouth Gesture. Just like the video screen, the tiers of informant B are on the left, those for informant A on the right side. It is not very often that the moderator interacts. Therefore, we skipped the total perspective in the online view and added a seventh tier for a summary of the moderator's interaction. To keep the tiers apart, they have different background colours for informants and moderator. A link allows switching to the German version (with translation and glosses in German).

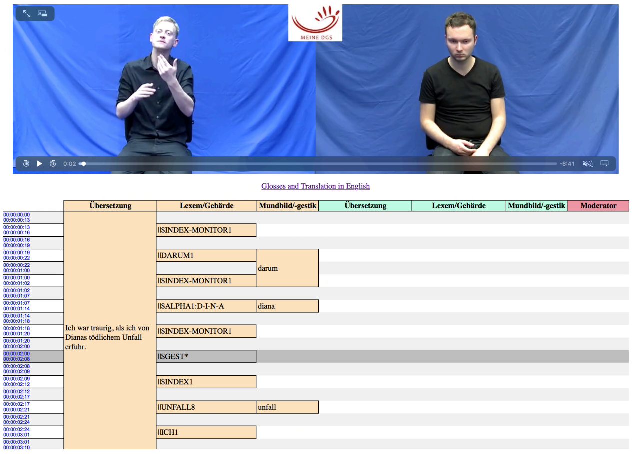


Figure 3 : Research Portal Annotation Tiers

3.2.1 Annotation Tiers

The German translation should be as close to the DGS utterance as possible. We did not aim for a free translation, because the translation should guide the mostly hearing student annotators. Contracted sign language interpreters conducted a first translation. The student co-workers splitted and time-aligned these texts into 'sentence'-like utterances. As Johnston (2016: 14) posits, these "translation sentences are not attempts to segment the [DGS] text into its potential language-specific syntactic or grammatical units". They are searchable and define preliminary utterance units when looking for the context of a sign token. The translation into English is a free translation. Its purpose is to give access to the content of the DGS videos to those knowing neither DGS nor German.

Mouthings are very frequent in DGS. They are an important clue to the meaning of a DGS sign token which, in combination with the sign form, can be used to search for the appropriate type the token should be matched to. Thus, we decided to also annotate mouthings in the phase of basic annotation. Mouthings are annotated in lower case to make them distinct from German words. As we focussed on the meaning of the mouthed word and not its actual articulation, at least the intended word (word stem) to be lip-read should be annotated. Incomplete mouthings are supplemented (in curly brackets), uncertainties are

[1] For performance reasons the videos have a framerate of 25 fps, the timeline instead follows a 50 fps rate to be consistent with the timecodes in the ELAN and iLex import files. As a consequence, the videos in the online view are not suitable for frame-to-frame inspection. For this, one has to use the download files.

marked by "??". As mouthings in DGS refer to German words, the articulation features are different from e.g. mouthed English words. We therefore do not provide a translation of mouthings.
Mouth gestures are movements of the mouth region with no connection to words of the vocal language. With a focus on lexical signs, we did not aim for classifying mouth gestures by form features. They are annotated in a simplistic way by just adding "[MG]" in the Mouthing/Mouth Gesture tier.
The annotation files are complemented by two HamNoSys tiers with notations of the citation form of types in the Lexeme/Sign tiers that are available after download. Annotation Conventions for the Lexeme/Sign tiers are explained in the following.

3.2.2 Annotation Conventions

There are two main aspects in which our approach differs from those of other sign language corpus projects: the role of mouthings which led us to implement a type hierarchy (double glossing) in the database model, and double-token tags in the token tier instead of separate gloss tiers for left and right hand.

3.2.2.1 Type hierarchy (double glossing)

In brief, we are convinced that following the principle of idiomaticity does not fit the needs of an adequate description of a sign language lexicon. The reason why (lexical) signs can cover a far wider range of meanings than words is iconicity. Sign languages exploit the possibilities to express the visually perceivable world in a visual-gestural modality which also allows for integrating words of the surrounding vocal language by way of mouthings. Conventionalisation should not only be applied for distinguishing lexical from productive signs, but also for sign-mouthing combinations (for further details see König et al. 2008, 2010, Konrad et al. 2012).
Glosses in the "Lexeme/Sign" tier refer either to a type or a subtype. Types correspond to lexical entries which have at least on conventionalised meaning. In order to group these form-meaning combinations, often expressed by conventionalised sign-mouthing combinations, we use subtypes. Each type (parent) has at least on subtype (child). Tokens of conventional sign-mouthing combinations are matched to the appropriate subtype, tokens of productive sign-mouthing combinations are matched to the type. This kind of pre-sorting supports the lexical description of sign types.
Glosses are labels for sign types/subtypes representing unique type entities in the lexical database and can be taken as ID-glosses, regardless whether in German or English (Johnston 2008; Konrad/Langer 2009). Glosses at the type level are marked by a superscript after the gloss name as e.g. FLACH1^ (PLANE1^). One of its subtypes is TISCH1 (TABLE1), without superscript. In iLex we annotate form deviation to the token tag and sort tokens for morpho-syntactic patterns or modification by using qualified types (see Konrad et al. 2012). In the Research Portal we only show types and subtypes. Tokens that differ from the types citation form are marked by an asterisk after the gloss name, e.g. TABLE1*.

The online view of transcripts not only allows to browse the annotations, but also can be used to list all tokens of a type and subtype lemmatised in the whole public corpus. By clicking on the gloss name in the "Lexeme/Sign" tier a new page opens with all the tokens that are matched to the corresponding type and/or subtype, irrespective whether the type or subtype gloss is clicked. In addition to the gloss name, several metadata are provided: region, format, age group, and sex. The following screenshot shows the tokens matched to the type SOUL2^ and the subtype EMBARRASING2:

SOUL2^

SOUL2^* Bremen | Discussion | 31-45f

EMBARRASSING2

EMBARRASSING2 Stuttgart | Subject Areas | 31-45f

EMBARRASSING2 Stuttgart | Deaf Events | 31-45f

EMBARRASSING2* Frankfurt | Experience of Deaf Individuals | 18-30f

Figure 4 : Listing of Tokens from Types and Subtypes

3.2.2.2 Double Tokens

Many researchers using e.g. ELAN as annotation tool have two token tiers, one for each hand. Two-handed signs are lemmatised by annotating the same gloss in each tier. In order to make the annotation easier and less time-consuming we opted for one token tier which allows for annotating one type for each hand. Two-handed signs are either annotated in the right or left hand slot: For asymmetric signs the slot of the active hand is used. For symmetric signs the right hand slot is used as a default.
A sign articulated with the right hand – being either a one- or two-handed sign – is displayed in the type-/subtype-gloss tier by one gloss. If the sign is articulated with the left hand, the gloss is preceded by a double bar, e.g. ||HAUS1A (HOUSE1A). A complex sign construction shows two glosses separated by a double bar, e.g. OMA2||$INDEX1 (GRANDMA1||$INDEX).

3.2.2.3 Glossing conventions

Although basic annotation of sign language texts should be as theory-neutral as possible, it cannot do without any theoretical assumptions. One is the distinction of three sign categories: lexical signs (cf. Johnston 2016: fully-lexical signs), productive signs (cf. Johnston 2016: partly-lexical signs), and others (cf. Johnston 2016: non-lexical signs). In the following we just mention some of the glossing conventions, for a detailed description see "Annotation Conventions" in the Research Portal.
Lexical signs are glossed by German (English) words. Different numbers are used to group lexical variants, e.g. FRAU4 (WOMAN4) and FRAU5 (WOMAN5). Phonological variants are grouped together by using the same gloss name and number followed by different letters, e.g. FRAU2A (WOMAN2A) and FRAU2B (WOMAN2B). Productive signs are glossed as $MAN (abbreviation for "manual activity"; $PROD for "productive sign"). For

grouping together type categories in a sorted type list, we use prefixes like $NAME- (name signes), $INDEX (pointing signs), $ALPHA (fingerspelling), or $GEST (gestures).

3.3 ANNIS

The platform-independent open-source search and visualization tool ANNIS comes along as both a web-application and a local version. ANNIS was put forth by a DFG project, the SFB632 "Information Structure: The Linguistic Means for Structuring Utterances, Sentences and Texts", realised by researchers of the University of Potsdam, the Humboldt-University of Berlin and the Free University of Berlin. While the project ended in 2015, ANNIS has been used by further projects ever since. It is meant to be a storage and search possibility for complex corpora with multiple layers that can originate from different annotation tools. Along with the growing number of multimodal corpora, ANNIS allows to implement video data as well as linking parts of a video with the associated annotations. It also enables users to directly search for annotations with the ANNIS query language (AQL; for more information see Rosenfeld 2010), that provides powerful search options. With every corpus that is published in ANNIS, search examples in AQL are provided, either automatically or preset by the researcher. Clicking on these example queries leads to their results. AQL allows searching for, inter alia, entries and metadata, sequences and hierarchical orders. Complex searches can be formulated in AQL, too, in accordance with the following scheme. First, one or more attribute-value pairs are defined. Second, the relationships between the nodes are defined, using among others the following operators: in-/direct precedence, in-/direct neighbourhood, in-/direct dominance and (identical) overlaps. Regular expressions can be used, too. All values can be negated and so can metadata. Search results can be displayed in different views, like syntax trees or dependency relation schemes. Since annotation tiers are not hierarchically linked in the Public DGS Corpus, results are presented in a KWIC (key word in context) table view, called grid. The size of the context is preset to five tokens both left and right of the search result (but can be varied). Once a search is successfully carried out, a frequency analysis on the search results can be conducted. For further statistical or other analyses, results can be downloaded in various formats. Results of a search or a frequency analysis can be shared via a link.

ANNIS was choosen as a third presentation format in order to enable users to directly search the data online without the need to register, download data, install new programs and learn a completely new query language. The essential features of the ANNIS query language might be familiar to most researchers engaged in corpus based research. Since many researchers might already be used to ANNIS or similar corpus search tools, the Public DGS Corpus in ANNIS is therefore mainly meant to address those researchers. Nevertheless, also users without a (corpus) linguistic background can find an easy access-point to a scientifically motivated approach to the data with ANNIS.

Using ANNIS requires getting familiar with the tool and its query language. Also, ANNIS is not meant to be another content-related access point. Users should at least roughly know the content, the metadata and the annotation conventions used. Watching the complete video collection in ANNIS will most likely be uncomfortable – for this matter, meine-dgs.de is more advisable. While ANNIS also provides the possibility to store corpora in a restricted area, to which only researchers are granted access after registration with a university e-mail address, the Public DGS Corpus will be released in the public area, in which no prior registration is needed.

The Public DGS Corpus is presented in ANNIS as follows. Both in the online and the local version, the ANNIS main page contains two stationary elements, namely a box where AQL queries can be typed in and a list of publicly accessible corpora. (On a fixed tab, a help page and a tutorial can be opened at any time). Each corpus is listed with a small icon leading to metadata information about the corpus.

Corpus information for DGS-Korpus (ID: 2813)

Metadata

Select corpus/document: DGS-Korpus

Name	Value
Contact	info@dgs-korpus.de
Project	DGS-Korpus
Project_Description	The DGS-Korpus project is a long-term project of the Academy of Sciences in Hamburg for the documentation of and research on German Sign Language (DGS). The aim is to collect sign language texts from Deaf people and to present parts of them as a public corpus, which will contain about 50 hours. The public corpus contains almost 400 episodes covering 18 different elicitation settings ranging from experience reports of Deaf individuals as well as discussions to story retellings and jokes. The data cover 327 informants from all over Germany and is meant to be representative for the everyday language of Deaf people throughout Germany.
Survey_Period	2010 - 2012
Website	http://www.dgs-korpus.de

Figure 5 : Public DGS Corpus Metadata in ANNIS

Metadata can also be added for individual documents. Thus, the document metadata can be used as values in queries.

Corpus information for DGS-Korpus (ID: 2813)

Metadata

Select corpus/document: BER01

Name	Value
Format	Experience Report
L_Age_Group	18-30
L_Gender	male
R_Age_Group	18-30
R_Gender	male
Recording_Date	2011-08-06
Region	Berlin (Berlin, Brandenburg, parts of Saxony-Anhalt)
Topics	Society, Traffic
annis:doc	BER01

Figure 6 : Document Metadata in ANNIS

Selecting the Public DGS Corpus leads to a list of example queries. Clicking on an example query leads to the search result. This is a user-friendly access to the data and gives a good first impression of the query language.
Search results are presented by means of two grids (for English and German annotations) that can be folded up and out at will. The video is displayed above. The grids contain the same tiers that are displayed on the Research Portal online view, namely three tiers per informant (Translation, Lexeme/Sign and Mouthing/Mouth Gesture) and one for the moderator.

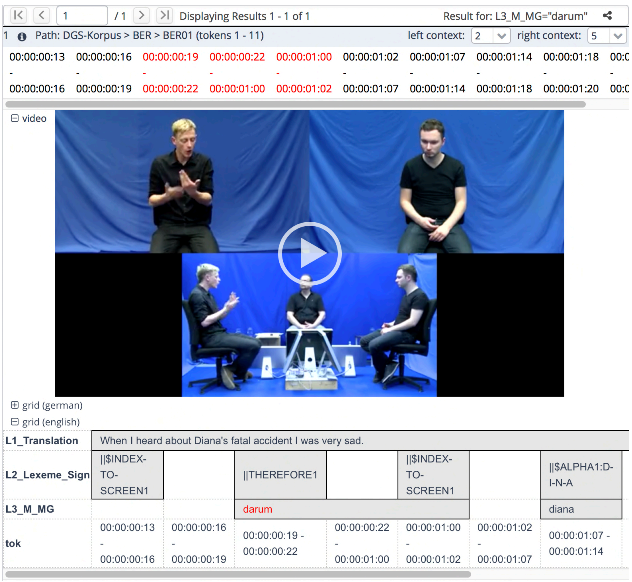


Figure 7 : Presentation of Results in ANNIS

With the presentation of the Public DGS Corpus in ANNIS, we provide access for corpus-based research that is based on a variety of different information such as different layers of annotation and relations between annotation tiers and metadata information. The publication of the public corpus in ANNIS makes the data scientifically usable, facilitates perusing of the data and allows the sharing of search results with other interested parties by means of a handy link.

4. Involving the Language Community

meine-dgs.de is a low-threshold website that provides an easy access point. For users from the Language Community, whose first language is a visual language and who therefore might feel more natural with signed information, we provide information about the low-threshold format meine-dgs.de by means of a signed video introduction. Also, meine-dgs.de is designed to be intuitively usable. It is not text-intensive, clearly structured, and in general mainly visually oriented, with clickable pictures and short access paths.
Furthermore, we included features that facilitate interactivity and the involvement of the Language Community, namely the "Mistake" button and the share function.

As for the "mistake" button, interactivity makes the data and its use even more attractive and helps to improve the quality of the published data. As described above, published data has gone through a process of reviewing and examining. Nevertheless, mistakes can never be completely avoided.
An interactive exchange and the establishment of a discussion about the intrinsic value of the data for specific use cases could increase the users' interest in the data. Members of the focus group, a group of informants that are well rooted in the Deaf Community, supported this idea. Although the increase of interactivity through a comment function would be a great advantage, we are also well aware that it is difficult to filter comments and sort out offensive, nonsensical or other undesired comments. While this is usually a typical and tolerated sideeffect of online platforms with comment functions, we aim to strictly avoid situations, in which anonymous users criticise or insult informants. The benefits and costs of a moderator-controlled platform must be weighed. Up to now, it is only possible to share videos from meine-dgs.de to other platforms and social networks. This allows users to draw attention to videos they find especially interesting or valuable and also enables users to get in contact with each other.
For the moment, we take this as a sufficient solution to initiate the building of a community, which will be observed and inspected from time to time, in order to detect a good moment to organise an interaction directly on meine-dgs.de.

5. Conditions of Use

Obviously, publishing video data and at the same time protecting the rights of the informants is more difficult than publishing data collections that consist of texts or audio files. The privacy of the informants themselves as well as all persons mentioned in the dialogues has to be respected. Since the sign language community is a relatively small community, small hints on the identity of third persons mentioned might be enough for identification. For these reasons, we exclude data from publication when in doubt and restrict the publication of metadata to very rough categories, age group, sex, and larger geographic region.
We also attach great importance to matters of ethics and therefore follow the wishes of the informants how their data can be used. As they need to cover all data in the public corpus, the licenses for using the data are therefore more restrictive than for some other projects. More permissive licenses are available only upon request.

6. Acknowledgements

This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

7. Bibliographical References

Bleicken, Julian / Hanke, Thomas / Salden, Uta / Wagner, Sven. (2016). Using a Language Technology Infrastructure for German in order to Anonymize German Sign Language Corpus Data. In: Calzolari, Nicoletta et. al. (eds.): *LREC 2016 Proceedings. Tenth International Conference on Language Resources and Evaluation (LREC), May 23-28, 2016, Portorož, Slovenia.* ELRA. pp. 3303-3306. [URL: http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/LREC2016-419_Paper.pdf; last access 2018-02-23].

Crasborn, Onno A. / Mesch, Johanna / Waters, Dafydd / Nonhebel, Annika / Van der Kooij, Els / Woll, Benice / Bergman, Brita. (2007). Sharing sign language data online: Experiences from the ECHO project. In: *International journal of corpus linguistics*, *12*(4). pp. 535-562.

Hanke, Thomas. 2002: iLex. A tool for Sign Language Lexicography and Corpus Analysis. In: González Rodriguez, Manuel / Paz Suarez Araujo, Carmen (eds.): *Proceedings of the third International Conference on Language Resources and Evaluation. Las Palmas de Gran Canaria, Spain; Vol. III.* Paris: ELRA. pp. 923-926. [URL: http://www.lrec-conf.org/proceedings/lrec2002/pdf/330.pdf; last access 2018-02-23].

Hanke, Thomas / Storz, Jakob. (2008). iLex – A database tool for integrating sign language corpus linguistics and sign language lexicography. In: *LREC 2008 Workshop Proceedings. W 25: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora.* Paris: ELRA. pp. 64-67. [URL: http://www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf; last access 2018-02-23].

Hanke, Thomas / König, Lutz / Wagner, Sven / Matthes, Silke. (2010). DGS Corpus & Dicta-Sign: The Hamburg Studio Setup. In: *P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Martínez Ruiz, A. Schembri (2010, Eds.) Corpora and Sign Language Technologies. [Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages. 7th International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta.].* Paris. ELRA. pp. 106-109. [URL: http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/Hanke_et_al_2010_Studio.pdf; last access 2018-02-23].

Hanke, Thomas / Konrad, Reiner / Langer, Gabriele / Müller, Anke / Wähl, Sabrina. (2017). Detecting Regional and Age Variation in a Growing Corpus of DGS. Poster presented at the Workshop: *Corpus- based approaches to sign language linguistics: Into the second decade. July 24, 2017, Birmingham*. [URL: http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/DGS-Korpus_Poster_Birmingham2017_Variation.pdf; last access 2018-02-23].

Johnston, Trevor. (2016). Auslan Corpus Annotation Guidelines. November 2016 version. [Online resource; URL:http://www.academia.edu/29690332/Auslan_Corpus_Annotation_Guidelines_November_2016_revision].

Johnston, Trevor. 2008: Corpus linguistics and signed languages: no lemmata, no corpus. In: Crasborn, Onno / Efthimiou, Eleni / Hanke, Thomas / Thoutenhoofd, Ernst D. / Zwitserlood, Inge (eds.): *LREC 2008 Workshop Proceedings. W 25: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora.* Paris: ELRA. pp. 82-87. [URL: http://www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf; last access 2018-02-23]

König, Susanne / Konrad, Reiner / Gabriele Langer. (2008). What's in a sign? Theoretical lessons from practical sign language lexicography. In: Quer, Josep (ed.): *Signs of the time. Selected papers from TISLR 2004.* Hamburg: Signum. pp. 379-404.

König, Susanne / Konrad, Reiner / Langer, Gabriele / Nishio, Rie. (2010). How Much Top-Down and Bottom-Up do We Need to Build a Lemmatized Corpus? Poster presented at the Conference: *Theoretical Issues in Sign Language Research Conference (TISLR 10), Sept 30 - Oct 2, 2010, Purdue University, Indiana, USA.* [URL: http://www.sign-lang.uni-hamburg.de/dgs-korpus/tl_files/inhalt_pdf/PosterTISLRTranskription1_R12.pdf; last access 2018-02-23].

Konrad, Reiner / Langer, Gabriele. (2009). Synergies between transcription and lexical database building: The case of German Sign Language (DGS). In: Mahlberg, Michaela / González-Díaz, Victorina / Smith, Catherine (eds.): *Proceedings of the Corpus Linguistics Conference (CL2009), July 20-23, 2009, University of Liverpool, UK.* [URL: http://ucrel.lancs.ac.uk/publications/cl2009/346_FullPaper.doc; last access 2018-02-23].

Konrad, Reiner / Hanke, Thomas / König, Susanne / Langer, Gabriele / Matthes, Silke / Nishio, Rie / Regen, Anja. (2012). From form to function. A database approach to handle lexicon building and spotting token forms in sign languages. In: *O. Crasborn, E. Efthimiou, E. Fotinea, T. Hanke, J. Kristoffersen, J. Mesch (2012, Eds.) Interactions between Corpus and Lexicon [Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages. 8th International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey.]* Paris. ELRA. pp. 87–94. [URL: http://www.lrec-conf.org/proceedings/lrec2012/workshops/24.Proceedings_SignLanguage.pdf; last access 2018-02-23].

Krause, Thomas / Zeldes, Amir (2016). ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities*, 31(1). pp. 118-139. [URL: https://doi.org/10.1093/llc/fqu057].

Langer, Gabriele / Müller, Anke / Wähl, Sabrina. (2018). Queries and views in iLex to support corpus-based lexicographic work on German Sign Language (DGS). [this issue].

Nishio, Rie / Hong, Sung-Eun / König, Susanne / Konrad, Reiner / Langer, Gabriele / Hanke, Thomas / Rathmann, Christian. (2010). Elicitation methods in the DGS (German Sign Language) corpus project. In: *P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Martínez Ruiz, A. Schembri (2010, Eds.) Corpora and Sign Language Technologies. [Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages. 7th International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta.]* Paris. ELRA. pp. 178-185. [URL: http://www.lrec-conf.org/proceedings/lrec2010/workshops/W13.pdf; last access 2018-02-23].

Rosenfeld, Viktor. (2010). An implementation of the Annis 2 query language. Humboldt University Berlin.

Quotation in Russian Sign Language: a corpus study

Vadim Kimmelman, Evgeniia Khristoforova
University of Amsterdam & Russian State University for the Humanities, Russian State University for the Humanities
Amsterdam & Moscow, Moscow
{vadim.kimmelman, evkhristoforova}@gmail.com

Abstract

We studied how quotation is expressed in naturalistic discourse in Russian Sign Language (RSL). We studied a sub-corpus of the online corpus of RSL containing narratives by eleven signers from Moscow. We identified 341 instances of quotation, including reported speech and reported thoughts. We annotated syntactic, semantic, and prosodic properties of the found instances of quotation. We found out that quotative constructions in RSL have the same basic structure as similar constructions in other spoken and signed languages. Furthermore, similarly to quotation in other sign languages, quotation in RSL can be marked by head and/or body movement and change in eye gaze direction. However, all of these markers are clearly optional, and a considerable number of examples do not include any of these markers. Furthermore, we found that, judging by the behavior of indexicals, RSL narratives in our dataset have a very strong preference for using direct speech. We discuss theoretical implications of the RSL data to the theory of quotation in sign languages.

Keywords: role shift, quotation, Russian Sign Language, corpus research

1. Introduction

Quotation concerns the situation when the signer or speaker conveys somebody else's words, thought, or attitudes. Quotation has been an important topic in linguistic research, as well as in philosophy of language (Brendel et al., 2011).

Much research has also been devoted to quotation (and more broadly, to both constructed speech and constructed action) in sign languages (Lillo-Martin, 1995; Janzen, 2004; Quer, 2011; Herrmann and Steinbach, 2012; Lillo-Martin, 2012; Cormier et al., 2016; Schlenker, 2017). In this paper, we add novel data by looking at quotation in Russian Sign Language (RSL): quotation in this language has never been studied before. Furthermore, we use corpus data to study how quotation is expressed in RSL to give a more objective impression of the amount of variation.

When the topic of quotation or constructed speech is discussed in sign linguistic literature, it is usually done in connection to *role shift*, that is, a specific constellation of non-manual markers that are used to convey somebody's speech or actions. These markers typically include head and/or body turns, eye gaze change (looking away from the addressee), as well as emotional facial expressions attributed to the author of the quote and not the signer. However, while these non-manuals seem to be frequent markers of quotation (Herrmann and Steinbach, 2012) and constructed action, some authors also note that they are not obligatory (Janzen, 2004; Cormier et al., 2016).

There are thus at least two ways that one might approach studying quotation in a sign language. One way is to study the properties of quotative constructions involving role shift, thus defining the construction in question both functionally and formally. Another way is to study the properties of all quotational devices, both with and without non-manual marking, thus only using the functional definition. We consider the latter approach as more prudent in studying quotation in RSL for two reasons. Firstly, since quotation in RSL has never been studied before, we cannot a priori presume that this language also uses role shift for quotation. Secondly, and more importantly, we think that the properties of quotation with role shift and the nature of role shift itself—which is an issue of many debates (see below)—can only be studied in the context of other quotational devices (if such are available in a sign language).

When studying quotation in sign languages, several theoretical questions are usually raised. One important point discussed by many researchers was the question of whether quotations marked by role shift are direct or indirect speech. Several researchers have argued against the intuitively appealing idea that quotation marked with role shift is direct speech. For instance, Lillo-Martin (1995) provided a number of arguments against analyzing role shift in American Sign Language (ASL) as a marker of direct speech, including evidence of syntactic subordination of the quote (but see also Lee et al. (1997) for arguments against this position), and the fact that role shift is not only used for constructed speech/dialogue but also for constructed action. In some sign languages, such as German and Catalan Sign Languages, quotation with role shift appears to show mixed behavior of some indexical elements, thus not conforming to the definition of either direct or indirect speech.

A related issue is the nature of the role shift itself (in both quotation and constructed action). Some researchers analyze it as a context-shifting device (Quer, 2011; Schlenker, 2017); some also point out similarities between role shift and agreement (Herrmann and Steinbach, 2012), while some analyze non-manual markers as demonstration, akin to emotional use of intonation and gestures in direct quotation in spoken languages (Davidson, 2015).

As will become clear, it is not possible to fully discuss these theoretical issues as applied to quotation in RSL due to usual limitations of corpus data. However, we will show that the properties of quotation in RSL that we observe in our corpus are at least indicative of certain approaches.

The paper is structured as follows. In section 2. we outline the methodology of our study. Section 3. contains the main results of the study. Finally, in section 4., we summarize the findings and discuss the consequences of RSL data for the general theoretical debates.

2. Methodology

We used a sub-corpus of the corpus of RSL (Burkova, 2015). We selected free narratives (personal stories) produced by 11 RSL signers from Moscow. We chose to investigate the narratives because this genre is most likely to contain quotation. Furthermore, we decided to only investigate signers from Moscow to avoid possible regional variation. However, this also means that our conclusions are only generalizable to the Moscow variant of RSL, and specifically to the narrative genre, if at all.

All signers in this sub-corpus live and work in Moscow, and this is also where the data has been collected in 2012. Three of the signers grew up in other regions and moved to Moscow as adults, so they might introduce some amount of regional variation. The signers include 7 females and 4 males, aged 30-58. 6 of the signers have deaf signing parents, but all have acquired RSL early in life.

The sub-corpus includes approximately 8000 signs (estimated by the number of glosses on the right hand tier) and 1200 sentences. Despite the modest size of the sub-corpus, it contains a large number of instances of quotation. We found and annotated 341 such instances, including 277 instances of quoted speech.

We annotated each instance of quotation in ELAN (Crasborn and Sloetjes, 2008) according to a number of features: (1) *Type of quote*: speech/thought/attitude; (2) *Predicate of quotation*: is there an overt predicate introducing quotation, and if so, which predicate? (3) *Author*: is the author of the quote the signer him/herself in the past or another person? (4) *Overt author*: is the author of the quote overtly mentioned? (5) *Non-manual markers*: eye gaze direction and body turns; (6) *Indexicals*: are there any indexical elements in the quote, and if so, is there reference shifted or non-shifted? (7) *Markers of subordination*: is the any evidence of syntactic subordination of the quote? (8) *Direct vs. indirect speech*: are there any signs of direct or indirect speech, such as the use of a complementizer?

We have also tried to annotate emotional facial expressions and head movements. However, this resulted in very low inter-rater reliability, so we did not analyze these annotations further. Furthermore, as Cormier et al. (2016), among others, have shown, emotional facial expressions are clearly not obligatory in quotation and also clearly occur outside of quotational contexts, and head movements also have a large number of functions unrelated to quotation. The question of how exactly these non-manuals interact with quotation in RSL thus awaits further research.

3. Results

3.1. Basic properties of quotation

Quotational constructions in spoken and signed languages have several constituents, some of them optional. An instance of quotation necessarily contains *the quote*, that is, the words or thoughts that are being quoted, and it can also contain *the introductory clause* which in turn consists of mentioning *the author* and *the predicate of quotation*, that is, a verb of speech or thought (1). In addition, some quotational constructions contain a marker of quotation, such as *like* in English (2).

(1) [She]$_{\text{author}}$ [said]$_{\text{predicate of quotation}}$: [“I’m so tired!”]$_{\text{quote}}$

(2) She was [like]$_{\text{marker of quotation}}$: I’m so tired!

This basic structure is clearly applicable to quotational constructions in RSL. Consider example (3): the sign IX-1 ‘I’ is the author, the sign SAY is the predicate of speech, and the rest of the clause is the quote. In addition, the quote can be marked by certain non-manuals which can be considered a marker of quotation; however, as we discuss below, the non-manuals are not obligatory and analyzing them as a marker of quotation might be unwarranted. Example (3) does not contain any head or body movement or eye gaze change that could be analyzed as marking quotation.

(3) IX-1 SAY IX-1 YES THROUGH MOSCOW TRAVEL
‘I said: Yes, I am traveling via Moscow.’[1] [2]

We analyzed all found instances of quotation with respect to the presence of these constituents. Similarly to spoken languages (Mathis and Yule, 1994), using a predicate of quotation and overtly mentioning the author of the quote are clearly optional in RSL. Table 1 summarizes the occurrences of overt reference to the author,[3] and Table 2 summarizes the occurrences of an overt predicate of quotation.

Overt author	No author	Constructed action
160 (47%)	118 (35%)	52 (15%)

Table 1: Overt author.

The third column in Table 1 refers to the cases when an instance of quotation (constructed speech) follows an instance of constructed action by the same author. In such case it would be redundant to use an overt sign to refer to the author again, so we separated this category.

Overt predicate	No predicate	Palms Up
96 (28%)	218 (64%)	27 (8%)

Table 2: Overt predicate of quotation.

The third column in Table 2 refers to the not infrequent cases when there is not predicate of the quotation, but the quotation is introduced by the Palms Up Gesture (McKee and Wallingford, 2011), as in example (4). While this happens often enough to be noticeable, it is important to emphasize that this gesture is multifunctional (McKee and Wallingford, 2011), which is also true for RSL. Therefore, we cannot be sure that it is used specifically as a marker of quotation, and does not have another unrelated function.

(4) IX-1 PU PLANE JUST ONE HOUR CL:FLY
‘I’m like: it’s just an hour by plane to get there.’[4]

[1] http://rsl.nstu.ru/data/view/id/358/t/176300/d/179150

[2] Each example is accompanied by the direct link to the on-line version of the corpus. Note, however, that (free) registration is required to access the data. We use standard glossing conventions in glossing the RSL examples. IX stands for index (a pointing sign); POSS - possessive; PU - Palms Up. Non-manual markers: eg - eye gaze, h - head, b - body, l - left, r - right.

[3] In this table the percentages do not add up to 100% due to the presence of a small number of unclear cases.

[4] http://rsl.nstu.ru/data/view/id/255/t/59280/d/61730

The most common verbs that are used as predicates of quotation are THINK, ASK, CALL, and TELL. The verb CALL is an interesting case as it is not a verb of speech or thought itself; instead it described the action of attracting someone's attention, but it is nevertheless often used to introduce a quote (5). This example also illustrates that this predicate can also introduce questions, and not only declaratives.

(5) CALL-A DAUGHTER POSS-2 PRESENT WHAT?
‘(She) asks: what should I give to your daughter?.’[5]

3.2. Non-manual markers

Similarly to other sign languages, quotative constructions in RSL are sometimes accompanied with non-manual markers, specifically with eye gaze change (looking away from the addressee), and body and/or head turns (6).

(6) IX-1 $\overline{\text{GOOD, FINISH}}^{\text{eg-l,h-l}}$
‘I (say): good, that's it.’[6]

However, all these non-manual markers are clearly optional. Moreover, their scope does not always align with the quote: sometimes only a part of the quote is marked non-manually, and sometimes the predicate or even the author is also marked with the same non-manuals. Table 3 summarizes the frequency of different scopes for eye gaze, Table 4 for body movements.

No marking	146 (43%)
Whole quote	47 (14%)
Part of quote	52 (15%)
Also predicate	38 (11%)
Also author	58 (17%)

Table 3: Scope of eye gaze.

No marking	166 (49%)
Whole quote	58 (17%)
Part of quote	63 (18%)
Also predicate	22 (7%)
Also author	32 (9%)

Table 4: Scope of body turns.

The large number of examples without eye gaze (146, 42%), and without body turns (166, 49%) show that these markers are clearly optional. Moreover, in 95 cases (26%), neither eye gaze nor body turns are used to mark the quote. Note that both reported speech and reported thought (and attitude) can occur with or without non-manual marking. Table 5 shows that the frequency of non-manual marking is similar for the two types of quotes, although reported thought seems to be marked more frequently than speech.
One might hypothesize that only direct speech is marked with non-manual markers in RSL, so for direct speech, these markers will turn out to be obligatory or nearly obligatory. We test this hypothesis in the next section.

[5]http://rsl.nstu.ru/data/view/id/366/t/136490/d/140020
[6]http://rsl.nstu.ru/data/view/id/257/t/57810/d/58940

Type of quote	*speech*	*thought*
Unmarked by eye gaze	122 (44%)	24 (37%)
Unmarked by body	146 (53%)	20 (31%)
Total	277	64

Table 5: Absent non-manual marking and quote type.

3.3. Direct vs. indirect speech

One of the main distinctions between direct and indirect speech is the behavior of indexical elements (Brendel et al., 2011), such as first and second person pronouns, time and place adverbs (*now, then*), and tense marking: in direct speech, such elements are interpreted with respect to the context of the quote itself (7), while in indirect speech the are interpreted in the context of the main utterance (8).

(7) John said to me yesterday: “I am tired now.” (I=John, now=yesterday, present tense=past interpretation)

(8) John said to me that I was tired then. (I=the speaker)

Therefore, we found all indexical elements in quotes in our data set, and annotated their reference. It turned out that the majority of examples (196, 57%) do not contain any indexicals. Furthermore, in a large number of examples, the author of the quote is the signer him/herself in the past (as in example (3)), so a first person pronoun refers to the signer irrespective of the context of interpretation. Such examples are ambiguous between direct and indirect speech.
Looking at examples with indexicals, the absolute majority (86 out of 91, 95%) contain indexicals interpreted in the context of the quote (shifted indexicals), so these examples can be characterized as direct speech.
For several sign languages, including German and Catalan, mixed behavior of indexicals has been reported (Quer, 2011; Herrmann and Steinbach, 2012). Specifically, while personal pronouns, such as IX-1 are interpreted as referring to the author and not the signer, an adverb like HERE within the same quote can be interpreted as referring to the situation of the main utterance. In our data set, we found two quotes with potentially mixed behavior of indexicals, but both instances involved multiple clauses which makes it possible to analyze them as sequences of direct and indirect quotes. We did not find any examples of a single clause with indexicals showing mixed behavior. However, the absence of such examples in our data set does not exclude the possibility that they are in fact grammatical. Further research is needed.
Since most examples with indexicals can be characterized as direct speech, we further studied non-manual markers in these examples. Contrary to the hypothesis mentioned in the previous section, these examples are not obligatorily marked with non-manual markers. Specifically, we found 31 examples (36% of all examples with shifted indexicals) not marked by eye gaze, 30 examples (35%) not marked by body movements, and even 15 examples (18%) not marked with any of these non-manual markers. We conclude that these non-manual markers are clearly not obligatory markers of direct speech.
We also discovered one clear marker of indirect speech in RSL, namely the complementizer THAT (9); we found 9

such examples produced by 4 signers. This complementizer is likely to be borrowed from Russian, as also indicated by the fact that in both languages it is homonymous with the question word meaning 'what'.

(9) eg-r,b-r
TELL-3 THAT IX-1 PU LAMP IX-1 NO
'I told her that my lamp was missing.'[7]

3.4. Syntactic subordination

For some sign languages, it has been claimed that quotes marked with role shift are syntacticaltly subordinate (that is, they are clausal arguments of the predicate of quotation (see e.g. Lillo-Martin (1995), but also Lee et al. (1997)).
Corpus data is not well suited to investigate this issue for RSL. We did find one clear piece of evidence that some of the quotes are subordinate clauses, namely the use of complementizer THAT (9): note that one cannot use this complementizer in a main clause. However, for the absolute majority of cases, we find no evidence of syntactic subordination of the quote. Specifically, we did not find clear cases of topicalization or wh-movement from the quote or center embedding of the quote. However, since such processes are not very frequent in general, we definitely cannot consider the absence of evidence here as evidence of ungrammaticality. Elicitation of acceptability judgments is necessary to further investigate this issue.

4. Discussion

In this study, we described basic properties of quotation in RSL based on narrative corpus data. We found out that quotative constructions have the same basic structure as similar constructions in other spoken and signed languages (3.1.). A somewhat surprising finding was that non-manual markers that can accompany the quote, while being similar to those described for other sign languages, are not obligatory and do not always align with the quote alone (3.2., compare for instance to Herrmann and Steinbach (2012)).
Another interesting finding is that, judging by the behavior of indexicals (and also by the use of the complementizer), RSL has a very strong preference for using direct speech, although indirect speech is also possible (3.3.). Note, however, that this can only be generalized to the genre we investigated, namely informal personal monologue narratives. It might be the case that, for instance, in more formal genres, more indirect speech would be used.
Finally, we found no clear examples of mixed behavior of indexicals (3.3.) and no evidence of subordination of quotes not containing the complementizer (3.4.). However, no strong conclusions about these issues can be based on corpus data alone, so they are left for future research.
The non-obligatory and not very systematic nature of non-manual markers which we observed does have some theoretical consequences. Specifically, it is difficult to analyze these non-manuals as context shift operators or as agreement markers (Quer, 2011; Herrmann and Steinbach, 2012), because there are clear examples of shifted indexicals in the absence of one or all of the markers.

[7]http://rsl.nstu.ru/data/view/id/259/t/89000/d/91051

We would argue that the nature of non-manual markers accompanying quotes in RSL is better captured by the demonstration theory proposed by Davidson (2015). Informally, non-manual markers are the signer demonstrating or re-creating certain behaviors of the author producing the quote. They are akin to emotional intonation and gestures that speakers of spoken languages can also use in quotation.

5. Acknowledgements

This research has been supported by the RSF (Russian Science Foundation, grant nr. 17-18-01184).

6. Bibliographical References

Elke Brendel, et al., editors. (2011). *Understanding quotation*. de Gruyter Mouton, Berlin.

Burkova, S. (2015). Russian Sign Language Corpus.

Cormier, K., Smith, S., and Sevcikova-Sehyr, Z. (2016). Rethinking constructed action. *Sign Language & Linguistics*, 18(2):167–204.

Crasborn, O. and Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pages 39–43. ELRA.

Davidson, K. (2015). Quotation, demonstration, and iconicity. *Linguistics and Philosophy*, 38(6):477–520.

Herrmann, A. and Steinbach, M. (2012). Quotation in sign languages: A visible context shift. In Isabelle Buchstaller et al., editors, *Quotatives. Cross-linguistic and Cross-disciplinary Perspectives*, pages 203–230. John Benjamins, Amsterdam.

Janzen, T. (2004). Space rotation, perspective shift, and verb morphology in ASL. *Cognitive Linguistics*, 15(2).

Lee, R. G., Neidle, C., MacLaughlin, D., Bahan, B., and Kegl, J. (1997). Role Shift in ASL: A Syntactic Look at Direct Speech. In Carol Neidle, et al., editors, *Syntactic structure and discource function: an examination of two constructions in American Sign Language*, pages 24–45. Boston.

Lillo-Martin, D. (1995). The Point of View Predicate in American Sign Language. In Karen Emmorey et al., editors, *Language, Gesture, and Space*, pages 155–170. Lawrence Erlbaum, Hillsdale, NJ.

Lillo-Martin, D. (2012). Utterance reports and constructed action. In Roland Pfau, et al., editors, *Sign Language: An International Handbook*, pages 365–387. Mouton de Gruyter.

Mathis, T. and Yule, G. (1994). Zero quotatives. *Discourse Processes*, 18(1):63–76.

McKee, R. and Wallingford, S. (2011). 'So, well, whatever': Discourse functions of *palm-up* in New Zealand Sign Language. *Sign Language & Linguistics*, 14(2):213–247.

Quer, J. (2011). Reporting and quoting in signed discourse. In Elke Brendel, et al., editors, *Understanding Quotation*, pages 277–302. De Gruyter Mouton, Berlin.

Schlenker, P. (2017). Super monsters I: Attitude and Action Role Shift in sign language. *Semantics and Pragmatics*, 10(9).

Where Methods Meet: Combining Corpus Data and Elicitation in Sign Language Research

Vadim Kimmelman, Ulrika Klomp, Marloes Oomen
University of Amsterdam
Amsterdam, the Netherlands
{v.kimmelman, u.klomp, m.oomen2}@uva.nl

Abstract

We discuss three case studies on various grammatical phenomena in Russian Sign Language (RSL) and Sign Language of the Netherlands (NGT) in order to compare corpus-based and elicitation-based approaches to sign linguistics. Firstly, we investigate impersonal reference in RSL using corpus search, informal elicitation, and an acceptability judgment task. Secondly, we examine argument structure and pro-drop licensing in NGT psych verb constructions using corpus search and a supplementary acceptability judgment task. Thirdly, we investigate conditional clauses in NGT based on corpus search, and contrast the findings with those from elicitation-based studies of conditional clauses in other sign languages. The three case studies highlight both the merits and limitations of combining different research methods as well as illustrate some of the issues that arise from doing so – and how they may be navigated. We conclude that corpus-based research serves to identify the boundaries of observed variation and describe both expected and unexpected patterns, while the underlying factors for these patterns can be investigated by eliciting data in more controlled contexts. Finally, we demonstrate that the differences in the results obtained via various research methods have important practical implications, in particular for sign language education.

Keywords: corpus research, elicitation, methodology, sign language

1. Introduction

The growing number of corpora of sign languages and the concomitant increase in corpus-based research in sign linguistics (Efthimiou et al., 2016) have made it important to evaluate corpus methods and compare them to other methods, specifically, elicitation.

For spoken languages, the advantages and disadvantages of corpus-based methods in comparison to elicitation have been described. Corpus data is more natural; it contains contexts; it shows a greater amount of variation. On the other hand, corpora by definition cannot provide negative data (if something is not attested in a corpus, it does not mean that it is ungrammatical), and some of the variation in the corpus might be due to performance errors (Hoffmann, 2006; Gilguin & Gries, 2009). It has therefore been suggested that combining the methods is a way to overcome the disadvantages of both (ibid).

However, for sign languages, no systematic methodological research to compare the various methods has been done so far. As a first step toward filling this gap, this paper discusses such methodological issues based on three case studies in two sign languages: Russian Sign Language (RSL) and Sign Language of the Netherlands (*Nederlandse Gebarentaal*, NGT).

In sections 2-4 we describe the case studies, and we summarize in section 5 how the different methodologies compare to and complement each other. Finally, we discuss the practical implications of the studies – in particular for sign language education – in section 6.

2. Case Study 1: impersonals in RSL

Many spoken languages have specialized impersonal pronouns that are used for impersonal reference (referring to humans but not specifying the referent exactly), such as *one* in English and *man* in German (Gast & Van der Auwera, 2013). We investigated how impersonal reference is expressed in RSL using a combination of corpus search, informal elicitation, and acceptability judgments (Kimmelman, in press).

Impersonal pronouns can be used in a variety of contexts, such as in existential contexts (*Someone has stolen my car*), universal contexts (*They eat snails in France*), and conditionals (*If you drink, you should not drive*). Furthermore, a language can use dedicated impersonal pronouns, such as *one*, but also use personal pronouns with impersonal reference, such as *you* and *they*. We aimed to find out which strategies (e.g., pronouns) are used in RSL and in which contexts they can be used.

2.1 Corpus Study

For the initial investigation, we used the RSL corpus (http://rsl.nstu.ru/) (Burkova, 2015). The corpus contains recordings of 43 signers of RSL from different regions; the data mainly consists of narratives (spontaneous or cartoon retellings), and some dialogues. The corpus has been glossed (separate tiers for the right and left hands), and sentence translations are also present. Since no special annotation for impersonal reference was provided, we searched for impersonal contexts indirectly. Specifically, we searched for the Russian words *kto-to* 'someone' and *kto* 'who', plural marking on verbs, and second person pronouns and verb forms as they can all be used in impersonal contexts in Russian.

It turned out that some constructions for impersonal reference could indeed be identified in the corpus. For instance, pro-drop (1a) and the indefinite pronoun SOMEONE (1b) can be used in impersonal contexts.

(1) a. BUS COME / SPEAK NUMBER [video]
'The bus came and they announced its number.'
b. SOMEONE PORTER MAYBE THROW.OUT [video]
'Someone – maybe a porter – threw him out.'

However, the severe size limitations of the RSL corpus – the total number of signs in the corpus as estimated by the number of glosses on the right hand is 25 000 – prohibited us from investigating all possible contexts and all strategies used for impersonal reference.

2.2 Informal Elicitation

To amend this, we also used informal elicitation, that is, a translation task with four native signers of RSL. We used a questionnaire from Barberà & Cabredo Hofherr (in press) which includes all typical impersonal contexts. The signers were presented with a context implying impersonal reference and were asked to translate a sentence that could be used in this context.

Using this method, we found a variety of means to express impersonal reference in RSL. Pro-drop is used in all impersonal contexts; the pronoun SOMEONE is used in existential impersonal contexts and in conditionals. In addition, we found that the plural pronoun IX-PL can be used by some signers in universal impersonal contexts (2).

(2) IX-PL SAY IX-A DRINK A.LOT
'They say he drinks a lot.'

So far, using corpus data and elicitation tasks turned out to deliver partially overlapping results (the use of pro-drop and SOMEONE was found in both types of data) and partially complementary results (some of the contexts and the use of IX-PL were only found through elicitation). However, on one issue, the corpus and elicitation delivered partially contradicting results. We wondered whether the second person pronoun IX-2 could be used impersonally in RSL. While corpus data provided some such examples (3) [video], the four signers unanimously claimed that this pronoun only had a personal reading. Surprisingly, two of them spontaneously produced an impersonal sentence with this pronoun before they were asked about it explicitly.

(3) ORDER IF SMOKE FUT IGNORE, IX-2 FINE 1500-3000
'If you smoke, you will be fined 1500-3000 rubles.'

2.3 Acceptability Judgment Task

In order to further investigate the latter issue, we conducted an experimental acceptability judgment task in which 16 RSL signers were asked to rate a variety of signed RSL sentences in context on a 5-point scale. We created stimuli in which the IX-2 pronoun could only be interpreted as impersonal (e.g. *You should not smoke if you are pregnant* explained to a man). Judgments showed a great deal of variation in their judgment, but participants generally did not consider the impersonal use of this pronoun ungrammatical.

2.4 Discussion

How can we explain the conflicting results of the corpus search, informal elicitation, and formal experimentation? We suspect that the variation can be attributed to borrowing of the impersonal use of the second person pronoun from (spoken/written) Russian. While signers use this pronoun impersonally (perhaps as a form of code-switching to Signed Russian) and are not opposed to it in a judgment task, they consider this construction to be too much Russian-like when asked about it directly.

This case study thus shows several things. Firstly, corpora are limited – and sign language corpora are especially small. Therefore, many grammatical phenomena cannot be studied in detail using corpus data alone. Secondly, combining corpus and elicitation data is often productive. Thirdly, it is possible that corpus data and (informal) elicitation will produce conflicting results, which reflects the previously attested tendency of signers/speakers to be stricter in judgments than in actual use (Labov, 1975), which emphasizes the importance of using corpus data to get a realistic view of variation.

In connection to the latter point, sign language data presents an additional problem due to the possibility of code-switching and general interaction with spoken languages/manual systems. It is clear that instances of code-switching and borrowing are present in corpus data (Bank, 2015), but signers aware of the differences between sign language proper and a manual system are likely to reject constructions which resemble those used in the spoken language. However, these direct judgments might not reflect the actual use of native signers and they do not necessarily distinguish code-switching from borrowing. Again, it is therefore useful to combine corpus data with elicitation.

3. Case Study 2: Psych-Verbs in NGT

In a study of the argument structure of NGT psych-verbs – verbs denoting a psychological state or the bringing about of a change in a psychological state (Levin, 1993) – a combination of corpus data (3.1) and acceptability judgments (3.2) were analyzed (Oomen, 2017). Challenges that arose as a result of combining the two methods are discussed in 3.3.

3.1 Corpus Study

We analyzed 309 annotated dialogues from the Corpus NGT (Crasborn et al., 2008), 181 clauses containing 37 distinct psych-verb forms were identified with the use of search terms. Analysis of these clauses revealed the following patterns:

- All lexical forms of psych-verbs except two are iconically body-anchored, i.e., they iconically refer to an aspect of the internal experience or external expression of a psychological state;
- Most psych-verbs, such as ANGRY, SAD, and WORRY, typically select just an Experiencer argument, with a Theme argument occurring only occasionally (29/159 examples); they are labeled type-A verbs. Three lexical forms (LOVE, HATE, MISS), labeled type-B verbs, require both an Experiencer and a Theme;
- The Experiencer occurs in subject position, while the Theme, if present, occurs in object position. Object Experiencer constructions were not attested, although two seemingly idiomatic periphrastic constructions with MAKE and a psych-verb were found;
- The Experiencer can be left non-overt, but apparently under the condition that its referent is first person: 17 such examples occurred (4a), as opposed to just one example with a non-overt non-first person referent;
- The restriction for non-first person referents does not hold when role shift markers are present: 27 examples with role shift and non-overt non-first person subject Experiencers (4b) were attested;
- The auxiliary glossed as AUX-OP (Bos, 1994), which agrees with subject and object through path movement modification, can co-occur both with type-A and type-B psych-verbs, although only two such examples were found for both types.

(4) a. ANGRY
'I am angry.'

rs

b. MOTHER COME / ANGRY
'My mother came. She was angry.'

Without going into detail, we integrate the findings into a theoretical account by proposing that iconically-motivated body-anchoring of psych-verbs triggers an association with first person, which (i) forces the Experiencer in subject position, and (ii) leads to a default first person interpretation of the Experiencer in the case of a non-overt argument. For further details, see Oomen (2017).

The theoretical analysis can only be read as a set of hypotheses, since the (un)grammaticality of constructions that are not attested in the corpus data cannot be proven on the basis of corpus data alone. We designed a small acceptability judgment task to surmount this problem.

3.2 Acceptability Judgment Task

With the acceptability judgment task, in which three native deaf signers of NGT participated, we aimed to test the following three hypotheses:

I. A Theme argument can be added to psych-verbs that typically select just an Experiencer (type A).
II. The directional auxiliary AUX-OP can co-occur with type-A psych-verbs, despite the fact that these verbs preferentially occur with just one argument.
III. A periphrastic object Experiencer construction with MAKE and a psych-verb is grammatical in NGT.

Testing a fourth prediction, namely that a subject Experiencer can only be dropped when it is a first-person argument, turned out to be infeasible (see 3.3 for details).

The hypotheses were tested with 11 sentence pairs that consisted of a scene-setting sentence introducing the relevant referents and a target sentence, recorded with a native signer of NGT. Examples of target sentences with intended translations for the respective hypotheses are given in (5). Note that (5b) actually includes two target sentences, which differ in the directionality of AUX-OP (as indicated by the subscripts). Both were included to verify that the auxiliary successively picks out the Experiencer and Theme – which would be the expected order if these thematic roles map onto subject and object, respectively.

(5) a. INDEX_1 INDEX_3 ANGRY
'I am angry with him.'

b. INDEX_{3a} AFRAID $_{3a/b}\text{AUX-OP}_{3b/a}$
'She_{3a} is afraid of it_{3b}.'

c. ACCIDENT INDEX_1 SAD MAKE
'The accident made me sad.'

Three signers were shown each of the sentence pairs in random order and were asked to make a acceptability judgment for each target sentence. In case of rejection of the sentence, they were asked to explain their choice and to provide an alternative.

The results were variable but largely confirmed expectations. Two signers judged sentences with type-A psych-verbs and two arguments (5a) grammatical while one signer did not, offering instead that the Theme argument be dropped or replaced by AUX-OP. In line with expectations, signers uniformly agreed that sentences including AUX-OP and a type-A psych-verb with an Experiencer-to-Theme trajectory are grammatical. Finally, sentences such as (5c) were rejected by two signers, but accepted by a third signer. All three signers offered a construction with a subject Experiencer and a psych-verb (e.g. INDEX_1 SAD) as an alternative.

3.3 Challenges

The results that emerged from the two data types – although similar – did not fully converge, which was mostly due to the variability in judgments in the experimental task. Admittedly, the number of participants was small, but this does not pre-empt the question how such variation should be interpreted. A larger pool of participants would not necessarily lead to an elimination of variation; many factors – both participant-related and task-related – make acceptability judgments "noisy, volatile, less objective, and less generalizable than previously assumed", as Gilquin and Gries (2009:3) point out. Acknowledging this is important when interpreting the results from a judgment task. For instance, the signer who rejected sentences with type-A psych-verbs and two arguments mentioned at a certain point that the stimulus or cause of an emotion (i.e. the Theme argument) is hardly ever relevant or important. We do not know whether this statement reflects a personal opinion, an attribute of NGT psych-constructions, or an artifact of the sentence pairs in that they somehow relegate the Theme argument to this status. Whatever the reason, it might have affected how she (and the other participants) judged the sentences, with conflicting results as a consequence. Nonetheless, the combination of corpus data – in which 29 examples with a type-A verb and two arguments were found – and elicited data makes a much stronger case for the grammaticality of such constructions.

Similarly, one signer accepted the periphrastic constructions with MAKE, while the other two signers unequivocally rejected them. The judgment of the first signer might be influenced by Dutch, which does allow such constructions – but one can think of a myriad of other explanations for the differences in judgments. Yet, again, while the construction has not been proven to be ungrammatical, the results from the corpus data – which only contained two such constructions – and the judgment task combined provide more convincing support that the construction must at least be very marked.

Thus, both these examples illustrate how combining corpus and experimental methods can facilitate the interpretation of results, even if they show subtle differences.

Analysis of the corpus data revealed an intricate interaction between use of role shift, grammatical person and overtness of the subject Experiencer in NGT sentences with body-anchored psych-verbs. Of the eight (2x2x2) possible combinations of values for each of these factors, one – constructions with a non-overt non-first person Experiencer without role shift – was basically unattested, which gives rise to the suspicion that it might be ungrammatical.

Confirmation from experimental data would be welcome. However, due to the number and nature of the variables involved, a dauntingly elaborate experimental set-up would be required. In order to properly test the interaction between each of the three factors in an acceptability judgment task, for instance, eight conditions would need to be included. Moreover, both subject drop and role shift are (far) more natural in longer stretches of discourse, which creates the need to design relatively long examples in order to avoid negative judgments for unintended reasons. This issue adds a significant extra layer of complexity to the matter.

Thus, the grammaticality of sentences with the described combination of factors is not easily refuted, which has implications for the way a theoretical analysis should be presented. On the other hand, precisely because of the number and nature of the variables involved, it seems plausible that the described pattern would not have been discovered had only controlled elicited data been used – which once again shows the merits of corpus research.

4. Case Study 3: Conditionals in NGT

Our third topic concerns an extensive corpus study into conditional clauses in Sign Language of the Netherlands (Klomp, in press). We compare our results to those from studies on conditionals in other sign languages, which have primarily been obtained through elicitation tasks.

4.1 Neutral and Counterfactual Conditionals

A typical example of a conditional clause in English is shown in (6) (in italics):

(6) *If it keeps on snowing*, I'll take the tram.

The first clause is called the conditional clause or antecedent and the second clause the main clause or consequent. From studies based on spoken languages, we know that some languages make a formal distinction between neutral conditionals (6) and counterfactual conditionals (7) (Dancygier, 1998):

(7) *If it had been snowing*, I would have taken the tram.

In (7), it is clear that the speaker knows that it did not snow; therefore, this type of conditional is called counterfactual. Dachkovsky (2008) observed that Israeli Sign Language (ISL) marks these two types differently non-manually. She found that neutral conditionals are marked by wide eyes, whereas counterfactual conditionals are marked by squinted eyes.

4.2 Conditionals in Sign Languages

Although research on conditionals in sign languages is limited, three general patterns have been described for all studied sign languages (e.g., Liddell (1986) on American Sign Language; Dachkovsky (2008) on ISL). Firstly, conditionals can be introduced by an (optional) manual marker, i.e., an *if*-conjunction. Secondly, conditionals are accompanied by raised eyebrows, and sometimes also by other non-manual markers. Thirdly, the antecedent precedes the consequent. On the basis of this, one might conclude that the similarities between sign languages are strikingly strong; however, it is important to note that almost all previous research on conditionals is based on elicited data.

4.3 Aim and Data of Current Case Study

The study aimed to describe conditional clauses in NGT and to make cross-linguistic comparisons. Furthermore, we were interested whether NGT marks neutral and counterfactual conditionals differently. Our data was extracted from the Corpus NGT (Crasborn et al., 2008) and consists of 407 CCs: 357 with a manual and (often) non-manual conditional marker and 50 with only non-manual marking. The former were found by searching on the gloss tier in ELAN, using the keywords IF (*als*) and SUPPOSE (*stel*). The latter were identified by searching for the Dutch conjunction 'if' (*als*) on the translation tier. The conditionals represent 51 signers from various regions in the Netherlands (age: 17-84).

4.4 Results

The data reveals that also in NGT, the antecedent precedes the consequent, and the antecedent can be marked both manually and non-manually. However, we observed considerable variation in this marking. Firstly, seven different signs were found that can function as an (optional) *if*-conjunction. Secondly, we found striking variation in the position of the eyebrows, indicating raised eyebrows are not an obligatory marker for conditionals in NGT at all times. Figure 1 shows that only a minority of the conditionals with a manual marker and a relatively small majority of conditionals without manual marker are accompanied by raised eyebrows. The difference between these two groups of sentences is significant: the eyebrows were less frequently raised (instead of furrowed or neutral) in sentences with manual marker (odds ratio = 0.34, p = 0.01, z = 2.46, 95% confidence intervals from 0.13 to 0.76).

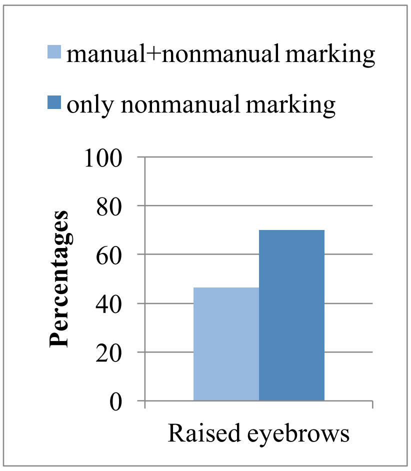


Figure 1: The percentages of conditionals with and without manual marker in which the eyebrows are raised.

Thirdly, we found that the use of other non-manuals, specifically the position of the head, varied as well. Neither this amount of variation in manual and non-manual marking, nor the optionality of raised brows when there is a manual marker has been described for conditionals in other sign languages. This raises the question if NGT conditionals are marked fundamentally differently from other sign languages, or if the different results are (partly) due to the different methodologies. Finally, we found few clear cases of counterfactual conditionals, suggesting that a different approach is needed to describe this category.

4.5 Discussion

Some of the variation that we found – particularly the variation in manual markers – can be explained by regional differences (Schermer, 2004). However, the amount of variation found in this study should likely be attributed to the methodology: it is based on more data from more signers from different regions, of more diverse ages and (language) backgrounds, than previous studies on conditionals in other sign languages. As we also described for the other case studies, this has advantages: we base our results on natural discourse and include many language users from different backgrounds. The disadvantages, on the other hand, are also clear: corpus data is not suitable for every aim (e.g., describing counterfactual conditionals) and results may be affected by independent factors (e.g., signers' background, context of the discourse), which we need to disentangle to interpret correctly (Hoffmann, 2006). Again, we conclude that elicited and corpus data complement each other.

5. Summary

The case studies we discussed show several clear examples of advantages and disadvantages of both corpus-based and elicitation-based methods in sign linguistics.

Two major limitations of corpus-based methods are the absence of negative evidence and size restrictions. The latter problem is unavoidable even with huge corpora of spoken languages, as some grammatical phenomena or lexical items might be too rare to be attested in even a very large corpus (Gilquin & Gries, 2009). However, for sign languages, this problem is even more drastic, as corpora of sign languages are relatively small. As such, even common phenomena, such as some impersonal constructions (case study 1) or counterfactual conditionals (case study 3), might be difficult to find.

There are also clear benefits of corpus-based methods in comparison to elicitation. A major advantage apparent in all three case studies is that corpus data better reflects the variation present in natural language use. However, as discussed in the third case study, the drawback is that the existing sign language corpora are not large or balanced enough to track the factors underlying this variation: we can observe the variation but not adequately explain it.

Another advantage of corpus data is the presence of extended contexts. Many phenomena, e.g., impersonals, pro-drop, and role shift, are only naturally used in longer stretches of discourse. Elicitation of constructed examples to describe such grammatical phenomena requires either the use of unnatural examples or longer test items, which makes elicitation tedious, as illustrated by case study 2.

A way to overcome the disadvantages of both elicitation and corpus-based methods is to combine the two (case studies 1 and 2) (Hoffmann, 2006). Thus, it is possible to identify the boundaries of the observed variation and describe both expected and unexpected patterns in the corpus – and then target the possible underlying factors by eliciting data in more controlled contexts.

Still, this is not a perfect solution, as corpus data and elicited data sometimes contradict each other (case study 1). We hypothesize that such contradiction often occurs due to more "puristic" judgments that signers (or speakers) give as compared to their natural language use (Labov, 1975). While this can happen with any type of grammatical phenomenon, this issue seems to be especially acute in the case of constructions borrowed from or influenced by a spoken language. Signers aware of the distinction between the manually-coded spoken language and the "real" national sign language tend to give negative judgments to constructions which resemble the ones used in spoken language, even if they themselves produce such constructions in naturalistic signing (case study 1). Neither corpus data nor elicited data (nor their combination) helps us unambiguously distinguish instances of borrowing from instances of code-switching.

Despite the limitations, we conclude that combining corpus and elicitation techniques is a strategy worth pursuing. With the increased availability of sign language corpora, we expect to see more studies in the near future delving deeper into the issues touched upon in this paper.

6. Implications for Sign Language Education

Here, we address the practical implications of results from studies combining corpus and elicitation methods by considering how they may be employed in the education of second language learners of a sign language. Our focus is on the Netherlands, specifically the University of Applied Sciences Utrecht (*Hogeschool Utrecht,* HU), where students are educated to become NGT interpreters or teachers. Clearly, these students need to reach fluency in NGT at a high level: at least level B2 in the Common European Framework of References for (Sign) Languages is required (Leeson et al., 2016).

The teaching method at the HU is, as much as possible, evidence-based (Van den Bogaerde & Boers-Visker, 2011). When teaching content is based on descriptions of NGT that are obtained either through corpus research or elicited data, there is a similar dichotomy as mentioned earlier. If teachers rely solely on signers' intuitions and results from elicited data, the teaching content would probably not reflect the variation that we encounter in the language. Conrad (2004) claims that ignoring the variation can even weaken the effect of teaching materials. It may lead, for instance, to students not learning structures that are commonly used among native signers. A concrete example: the most frequently found manual conditional marker in case study 3 is not included in the teaching materials of the HU – probably due to the fact that it is a regional variant.

On the other hand, if the teaching content is only based on corpus data, the input might be too varied. It is not the objective of the course to teach students all possible variants and dialects of NGT, since native signers also do not master all varieties. Furthermore, overreliance on corpus data in teaching may create the risk of implementing ungrammatical structures in teaching materials (performance errors, Hoffmann, 2006). This is indeed a paradox: while the aim is for students to master a language at a high level by getting natural and varied input, they at the same time benefit from clear rules and restricted variation. In the words of Aijmer (2010: 2): "(…) teachers (and learners) look for simple answers to grammatical problems in terms of what is right and wrong and shy away from the fuzzy picture of language as used in the corpus concordance."

Since more and more sign language corpora are becoming publicly available, the issue discussed above has become more relevant. We suggest that also in teaching materials, there are clear benefits of combining results of corpus data and elicited data. When there are clear rules, these can be offered to the students – but the variation that a language presents are best not ignored. When there is no corpus of a particular sign language available, one can think of cooperating with native signers to provide students with varied and qualitative input that reflects the variation in the language. For the Netherlands, the latest teaching methods are already focused on offering ample input and letting the students (initially) detect the rules themselves (Van den Bogaerde & Boers-Visker, 2011). When many different signers provide this input, the variation will likely be included naturally. Furthermore, if sign language teachers are aware of the extent of variation in different linguistic phenomena, they can keep that in mind when providing feedback on a student's language production. For linguists working with various methodologies, it is important that they are aware of the potential value of their results for language education and that they make an effort to make them accessible for language teachers.

7. Acknowledgements

The research discussed in this paper has been supported by Dutch Science Foundation (NWO) grant 360-70-520, and European Commission Horizon2020 grant 693349.

8. Bibliographical References

Aijmer, K. (2009). Introduction: Corpora and Language Teaching. In K. Aijmer (Ed.), *Corpora and Language Teaching.* Studies in Corpus Linguistics Vol. 33. Amsterdam/Philadelphia: John Benjamins Publishing Co, pp. 1-10.

Bank, R. 2015. *The ubiquity of mouthings in NGT*. Nijmegen: Radboud University Doctoral dissertation.

Barberà, G. & Cabredo Hofherr, P. (Eds.) (in press) *Sign Language &* Linguistics*: Special issue of Impersonal Reference in Sign Languages*.

Bos, H.F. (1994). An auxiliary verb in Sign Language of the Netherlands. In I. Ahlgren & B. Bergman (Eds.), *Papers from the First International Symposium on Sign Language Research*. Stockholm: Swedish Deaf Association, pp. 81-92.

Conrad, S. (2004). Corpus linguistics, language variation, and language teaching. In J. McH. Sinclair (Ed.), *How to Use Corpora in Language Teaching.* Studies in Corpus Linguistics Vol. 12. Amsterdam/Philadelphia: John Benjamins Publishing Co, pp. 67-85.

Dachkovsky, S. (2008). Facial expression as intonation in Israeli Sign Language: The case of neutral and counterfactual conditionals. *Signs of the time. Selected papers from TISLR 8*. Hamburg: Signum, pp. 61–82.

Dancygier, B. (1998). *Conditionals and prediction: Time, knowledge, and causation in conditional constructions.* Cambridge Studies in Linguistics 87. Cambridge: Cambridge University Press.

Efthimiou, E. et al. (2016, Eds.) *Corpus Mining* [Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages. 10th International Conference on Language Resources and Evaluation, LREC 2016, Portorož, Slovenia.] Paris: ELRA.

Gast, V. & J. Van der Auwera (2013) Towards a distributional typology of human impersonal pronouns, based on data from European languages. In D. Bakker & M. Haspelmath (eds.), *Languages Across Boundaries. Studies in Memory of Anna Siewierska*, 31–56. Berlin: Mouton de Gruyter.

Gilquin, G. and Gries, S.Th. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory*, 5(1):1-26.

Hoffmann, T. (2006). Corpora and introspection as corroborating evidence: The case of preposition placement in English relative clauses. *Corpus Linguistics and Linguistic Theory,* 2(2): 165-195.

Kimmelman, V. (in press). Impersonal reference in Russian Sign Language. *Sign Language & Linguistics.*

Klomp, U. (in press). Conditional clauses in Sign Language of the Netherlands: A corpus-based study. *Sign Language Studies*.

Labov, W. (1975). *What is a linguistic fact?* Lisse: Peter de Ridder.

Leeson, L., Van den Bogaerde, B., Rathmann, C., and Haug, T. (2016). *Sign Languages and the Common European Framework of Reference for Languages: Common Reference Level Descriptors.* Strasbourg-Cedex: Council of Europe Publishing.

Levin, B. (1993). English verb classes and alternations: A preliminary investigation. Chicago, IL: University of Chicago Press.

Liddell, S. (1986). Head thrust in ASL conditional marking. *Sign Language Studies,* 52: 244–262.

Oomen, M. (2017). Iconicity in argument structure: Psych-verbs in Sign Language of the Netherlands. *Sign Language & Linguistics*, *20*(1), 55-108.

Schermer, T. (2004). Lexical variation in Sign language of the Netherlands. Unpublished manuscript. https://testsiteen.gebarencentrum.nl/media/33555/92_file1.pdf. (downloaded 5 August, 2016).

Van den Bogaerde, B. and Boers-Visker, E. (2011). Theoretische onderbouwing met betrekking tot keuzes voor de modules Taalbeheersing Nederlandse Gebarentaal in het kader van de wijziging van het curriculum van de opleiding Leraar/Tolk Nederlandse Gebarentaal. Unpublished report.

9. Language Resource References

Burkova, S. (2015). Russian Sign Language Corpus. Retrieved January 1, 2018, from http://rsl.nstu.ru/

Crasborn, O., Zwitserlood I., and Ros, J. (2008). The Corpus NGT. A digital *open access* corpus of movies and annotations of Sign Language of the Netherlands. Centre for Language Studies, Radboud Universiteit Nijmegen. URL: http://hdl.handle.net/hdl:1839/00-0000-0000-0004-DF8E-6. ISLRN: 175-346-174-413-3.

What Corpus-Based Research on Negation in Auslan and PJM Tells Us About Building and Using Sign Language Corpora

Anna Kuder[1], Joanna Filipczak[1], Piotr Mostowski[1], Paweł Rutkowski[1], Trevor Johnston[2]
[1]Section for Sign Linguistics, Faculty of the Polish Studies, University of Warsaw,
Krakowskie Przedmieście 26/28, 00-927 Warsaw, Poland
[2]Department of Linguistics, Faculty of Human Sciences, Macquarie University, Sydney, NSW, Australia, 2109
{anna.kuder, j.filipczak, piotr.mostowski, p.rutkowski}@uw.edu.pl, trevor.johnston@mq.edu.au

Abstract

In this paper, we would like to discuss our current work on negation in Auslan (Australian Sign Language) and PJM (Polish Sign Language, *polski język migowy*) as an example of experience in using sign language corpus data for research purposes. We describe how we prepared the data for two detailed empirical studies, given similarities and differences between the Australian and Polish corpus projects. We present our findings on negation in both languages, which turn out to be surprisingly similar. At the same time, what the two corpus studies show seems to be quite different from many previous descriptions of sign language negation found in the literature. Some remarks on how to effectively plan and carry out the annotation process of sign language texts are outlined at the end of the present paper, as they might be helpful to other researchers working on designing a corpus.

Our work leads to two main conclusions: (1) in many cases, usage data may not be easily reconciled with intuitions and assumptions about how sign languages function and what their grammatical characteristics are like, (2) in order to obtain representative and reliable data from large-scale corpora one needs to plan and carry out the annotation process very thoroughly.

Keywords: sign language, corpus linguistics, corpus building, annotation, tagging, negation, non-manuals

1. Introduction

Representative and reliable data are indispensable in conducting linguistic research on sign languages. Due to significant sociolinguistic variation, resulting from numerous distinctive acquisition and usage patterns found in signing communities, researchers are often unable to draw clear generalizations concerning sign language grammars from individual signers' intuitions (as such judgments are not always accepted unanimously by other signers). The fact that sign language grammars have not been standardized to the extent typical for languages with a long tradition of writing and schooling (like English) comes as no surprise taking into account that Deaf people usually live dispersed within much larger speaking communities; sign languages are fairly young; and the inter-generational transmission of language in signing communities is often interrupted. To explore the extensive inter- and intra-signer variation, more and more research groups have decided to undertake the task of creating a corpus of the sign language they work on. Among those projects are: the Dutch Sign Language (NGT) corpus[1] (Crasborn and Zwitserlood, 2008), the British Sign Language (BSL) corpus[2] (Schembri et al., 2013) and the German Sign Language (DGS) corpus[3] (Hanke et al., 2010). More projects are underway. Basing linguistic analyses of the communication of the Deaf on real usage data (rather than on intuitions of individual signers) is becoming a methodological standard worldwide.

Our current work also belongs to the field of sign language corpus linguistics. In this paper, we would like to discuss our study on negation in Auslan (Australian Sign Language) and PJM (Polish Sign Language, *polski język migowy*) as an example of experience in using corpus data for research purposes (cf. Filipczak et al., 2015). The Auslan and PJM teams agreed upon fundamental methodological issues but actually worked separately on their own corpus material. Interestingly, both teams then made very similar observations about their annotation procedures and the phenomena they were revealing. These findings are outlined in the present paper as they might be of interest to other researchers working on corpus annotation and, in particular, on negation in sign languages.

2. Building and Annotating a Sign Language Corpus

Needless to say, building a sign language corpus is extremely time-consuming and labor-intensive. In most projects that are currently being developed, Deaf people are filmed in pairs as they respond to elicitation materials shown to them on a screen (see, e.g., Hanke et al., 2010; Rutkowski et al., 2017). Once videos are collected, they need to be annotated (Johnston, 2010). When starting the annotation process, it is vital to create written translations of as much of the recordings as possible as a matter of priority, even before glossing annotation starts. Translations are invaluable for being able to locate potentially interesting parts of the text in order to prioritize what should be glossed first. Translations can be prepared by Deaf signers, bilinguals, or hearing interpreters. It is important to employ a number of translators in order to have each chunk inspected by more than one person to ensure there is broad agreement. Individual sign glossing can be compromised if the overall meaning is not first established. (However, if there is unresolved disagreement among competent signers, this is also relevant and interesting. It may point to some real ambiguity or indeterminacy in the structure of the utterance that linguists need to take account of.)

When it comes to assigning glosses to individual signs one can either have a predefined lexical database or build the lexicon as one annotates the material. From our experience, either strategy will help ensure that the task is carried out consistently. Each lexeme needs to have its own unique label (assigned to every occurrence of the sign). Once glossed, the video material is machine readable and ready

[1] www.ru.nl/corpusngten/about-corpus-ngt/latest-news/
[2] www.bslcorpusproject.org/project-information/
[3] www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/the-project.html

to be used in linguistic research. (One must have confidence that the sign tokens identified during any searching, sorting and counting of the corpus are *all instances of the particular type* that one is interested in, as well as representing *all of the instances of that type* in the corpus.)

3. Negation study

The study reported in this paper is one of the first studies on sign language negation discussing corpus data. There is a corpus-based study by Oomen and Pfau (2017) concerning sentential negation in NGT. However, our work is the first one to compare negation data extracted from two independently created sign language corpora. It should be noted that there exists a widely-cited typology of negation patterns in sign languages (Zeshan, 2004; 2006), however, it was proposed on the basis of individual signers' grammaticality judgments and questionnaire data. Research based on corpus findings (for NGT, as well as for Auslan and PJM) offers a completely new perspective on Zeshan's typology.

3.1 Sources of Data

The source of data for the Australian negation study was the Auslan corpus – the first sign language corpus in the world. The Auslan archive[4] consists of 1100 video clips which, taken together, last approximately 300 hours. 100 Deaf signers were recorded for the purpose of creating the corpus; each of them performed 11 elicitation tasks during the recording session. Video recordings were edited and uploaded into the ELAN annotation software (Crasborn and Sloetjes, 2008). The Auslan corpus annotation is an ongoing process. So far, more than 350 clips have annotation files containing annotation at different levels of detail.

For the Polish negation study data were drawn from the PJM Corpus that is currently being compiled at the University of Warsaw by the Section for Sign Linguistics[5]. As of 2017, 134 Deaf informants were recorded. As each recording session lasts approximately 4-5 hours, the whole dataset exceeds 600 hours of raw HD video material. Obtained films were compressed and uploaded to the iLex software (Hanke and Storz, 2008), used for the purposes of annotation. Before being annotated, each video recording is segmented into more than 20 short video clips that correspond to elicitation tasks performed by the informants during the recording session. In the annotation process, the PJM Corpus team has so far identified over 5500 different lexemes (which have been divided into approximately 15,000 sublexemes), glossed approximately 505,000 individual sign tokens, translated more than 10,000 PJM clauses into Polish sentences and tagged approximately 100,000 tokens for their grammatical features. The annotation of the PJM Corpus is an ongoing process.

3.2 Data Annotation and Tagging

When analyzing negation in Auslan and PJM, we needed to be able to identify all manual signs associated with negation, as well as all occurrences of headshaking, a non-manual feature that is often interpreted as the marker of negation in sign languages (Zeshan, 2004; 2006; Pfau, 2015). Each team conducted two rounds of annotation/tagging specifically for the purposes of this study, on top of already existing annotations in each corpus. As the general annotation guidelines for the Auslan and PJM corpora are different, the negation tagging systems employed by the two teams also differed substantially, which makes the fact that the results were quite similar (as shown below) even more interesting.

For the Australian negation study, 413 video clips (24.7 hours of signed interaction) that had previously been segmented into signs and then glossed were examined. The annotation files for these clips were produced by 89 of the 100 individuals in the corpus. At the beginning of this study, approximately 9000 clauses had already been identified in these files during previous research. However, in only 89 of these was the entire text segmented into clauses and given time-aligned translation into written English. The remaining 324 clips already contained clause boundary annotations only at points that had been relevant to corpus-based research prior to this study. The 89 texts contained monologic spontaneous narratives, re-tells or elicited responses to visual stimuli (pictures and videos), or responses to interview questions involving dialogue with the interviewer and another participant also being interviewed. 375 of the 413 files had comprehensive time-aligned translations in written English and these accounted for 12 hours of recordings.

Taking into account that the Auslan data were prepared as outlined above, there were three ways in which it was possible to locate all instances of negation in them:

- searching the gloss annotations for all instances of Auslan signs known to be associated with negation or negative semantics, and investigating the relevant clauses for headshaking;
- searching the English translations for any words or word forms associated with negation in English and investigating the aligned Auslan clause or clauses for negative signs and/or headshaking;
- visually inspecting videos for all headshakes and annotating the co-occurring clause for the presence or absence of negation.

Each identified gloss that was negation-related was tagged for the presence or absence of head movement: headshake (HS), one strong turn of the head (HS1) wobbling (WOBBLE), tilting-back (TILT-BACK), or a side to side motion (SIDE-TO-SIDE). Any signs in the clause that did not display any of these movements was tagged as having no headshake (NHS) to clearly signal that the clause had been investigated for head movement, and to enable later searches for negation-related clauses that did not have a headshake in them. Headshakes that were observed to occur when no manual sign was being performed were also annotated over a place-holder gloss annotation on the glossing tier. Each identified clause was given a free translation and a literal or close translation into written English, if this had not already been done.

While applying the third tagging strategy – visually scanning the videos in search of the headshakes that did not co-occur with any negative manual sign – two phenomena became obvious: first, that headshake occurring during a manually-negated clause often seemed to make its own semantic contribution to the clause rather than just being another marker of negation; second, that nodding was also not only an extremely frequent head movement generally,

[4] https://elar.soas.ac.uk/Collection/MPI55247

[5] www.plm.uw.edu.pl/en

but even also occurred during manually-negated clauses. Consequently, the second round of annotation was conducted with more detailed tagging in order to identify when and why headshaking and nodding were used. In this round, negated clauses were tagged in such a way as to distinguish a number of pragmatic or semantic contexts in discourse by further specifying the grammatical class tag of the negation-related sign or by tagging added to the clause. A few different functions of negation-related signs were distinguished and tagged:

- response – if the clause within which the negation-related sign was found was in immediate response to a question from the interlocutor, or expressed a negative appraisal of what the interlocutor had just said;
- reprise – if temporally the sign was the second negation-related sign in a clause and appeared after the verb or another core constituent;
- imperative – used when inspected clauses were imperative;
- contrastive – used when inspected clauses presented an alternative.

With respect to the clause as a whole, two types of self-directed responses by the signer were identified: one to a topic and the other to a rhetorical question. The former were tagged as *clause internal responses* and the latter as *clause external responses*.

With respect to nodding, the head movement annotation of signs within negated clauses were changed from the default NHS for those that did not have a HS, to NOD if that is in fact the head movement that co-occurred with that sign.

In sum, all the manually-negated clauses and their associated head movements were identified in 413 ELAN files. Of these 89 files had all clauses and all headshaking behavior identified, irrespective of the presence or absence of negation. These 89 files comprised of 6327 clauses, of which 144 were negated. The number of clauses identified in the entire reference dataset had risen from approximately 9000 to 12,661 of which 1672 were tokens of clause negation[6].

The PJM team proceeded with their data in a slightly different manner. Note that the PJM Corpus material is generally glossed in the first step of the annotation process and translated and segmented into clauses on the basis of that. Since the negation study started before the segmentation and translation of the inspected data, it was not possible to search for cases of negation via Polish written translations in the whole dataset. In the first round of negation tagging, the Polish team focused solely on the third of the above-mentioned methods of locating cases of negation. The PJM Corpus was visually inspected for all occurrences of headshakes and negative signs, whether they co-occurred or not. Two tiers dedicated to this study were created in the iLex software and added to all transcripts. The NMNS_HEAD tier was used to tag all observed horizontal (left-to-right) head movements with respect to their role in the signed text. The following tags were used:

- hsh_NEG – when the observed headshake was associated with negation;
- hsh_ALT – when the occurring headshake was a marker of alternative;
- hsh_CL – meaning that the occurring headshake was part of the classifier (depicting sign) construction;
- hsh_Q – meaning that a left-to-right head movement was associated with a question, either produced with hands or purely non-manual;
- hsh_OTH – meaning that the observed headshake had a different function than any of the above.

When a sign associated with negation was produced with hands but no headshake was visible, the hsh_Ø tag was inserted into the NMNS_HEAD tier. Then, each of the identified headshakes was annotated with respect to its part-of-speech status on the second tier dedicated to negation (labeled as NEG_MAN). When the produced headshake was not associated with any manual sign or when it clearly did not target the manual sign it was co-articulated with, the Ø tag was used in the NEG_MAN tier. In total, 725 individual tasks (video clips) from the PJM Corpus were examined in this manner. Those clips were produced by 75 Deaf signers; they lasted approximately 103 hours in total and contained 244,000 individual sign tokens. Text types represented in the dataset included: re-tells of signed texts or visual stimuli, responses to visual stimuli, narratives and elicited, as well as free, conversations.

After the first stage of negation annotation was completed, the second round was conducted in order to specify the function of 'non-negating' headshakes. In this second round, 140 video clips, consisting of approximately 47,000 tokens, were inspected once again for all occurrences of negative headshakes that did not target any manual sign (i.e., the combinations of hsh_NEG and Ø tags). Those cases were marked with one of the following tags:

- neg_dec – when the headshake was used to change the polarity of the clause;
- neg_resp – when the headshake functioned as a marker of a question asked in the discourse (either by the interlocutor or the signer himself/herself);
- neg_imp – when the headshake functioned as an imperative marker;
- phatic – when the headshake was used only to show the signer's engagement in the discourse;
- meta-comment – when the headshake was a meta-comment to the narration built by the signer;
- discourse() – when headshake was an additional discourse marker; additional information was inserted in the brackets.

Simultaneously, all occurrences of morphologically-negated signs that were not accompanied by a headshake (namely all hsh_Ø + V_neg tag combinations) were inspected to assess whether different types of head movements (e.g. nodding, tilting-back or turning of the head) did not appear in such cases. When this was the case, the annotation was corrected accordingly.

3.3 Key Findings

In her seminal paper, Zeshan (2004) proposed a typology of sign languages with respect to negation. She studied 38 sign languages in her cross-linguistic survey and on the basis of this research distinguished two types of languages: manual-dominant sign languages that use mainly manual elements (negative particles and verbs articulated by hands)

[6] Note that this last figure does not represent an accurate guide to the proportion of negated to non-negated clauses in the corpus overall: negated clauses have been deliberately targeted as part of this study so their numbers are inflated.

to express negation, even though headshake could accompany these negative lexical signs; and non-manual-dominant sign languages in which negation is primarily conveyed by non-manual elements occurring during the production of the negated constituent, even though the non-manuals could also sometimes be accompanied by negative lexical signs.

Part of the present study aimed at classifying Auslan and PJM with respect to this typology. In order to perform this task, we analyzed and compared the data obtained during the annotation process.

We found that almost all (approximately 97%) of the grammatically negative clauses in the Auslan corpus included a negation-related sign and of these 61% overall also included a headshake during, at minimum, the production of that sign. In other words, only 3% were negated only non-manually. On these figures one would conclude that Auslan is an extreme manual-dominant sign language for negation: only a tiny fraction of negated clauses appear to use only headshake.

However, while scanning the data, more variation in the head movements accompanying negation was revealed. For example, nodding was observed to occur over the negation-related sign in 43 negated clauses. The role of this head movement was independent of the manual clause negation: it could not be construed as a negating element itself. Rather it appeared to reinforce the negation already present in the clause and/or expresses part of the signer's stance towards what the interlocutor has just signed or some discourse presupposition they both share. So the question arose: is it possible that in some manually-negated clauses headshake is also, like nodding, not part of the negation, but contributes additional information, albeit negative, about those grammatically-negated clauses?

In order to answer this question, headshaking in non-negative sentences was investigated. When we systematically scanned the subset of 89 recordings that had comprehensive annotations, we found almost 200 non-negated clauses with headshakes but only 5 clauses negated only by means of a headshake. This means there are 40 non-negating headshakes to every one headshake-only negator. These 89 files contained in total 145 negated clauses. Since 65% of manually-negated clauses were also accompanied by a headshake this means that this non-manual is associated with approximately 94 instances of clause negation, compared with approximately 250 instances where it is not. Clearly headshaking in Auslan is used more frequently outside of grammatical clause negation than within it.

The tagging of headshakes in negative environments with different discourse functions allows seeing the impact of the context on the likelihood that headshake will also be present during the production of a manually-negated clause. The co-articulation of headshake with apparently straightforward manual negation reduces from 65% to 50%. More telling, the rate of headshaking increases significantly (up to 89%) when the utterance is part of a response frame rather than merely the assertion of a negative state of affairs (which the manual negation is achieving anyway in virtually all Auslan negated clauses). The cases marked as 'contrastive' are accompanied by a headshake in 81% and 'reprises' in 62%. Negation was found to be fairly uncommon in imperative clauses.

As for PJM, even before starting the study presented in this paper, we were aware that the language does not fit easily in any of the types proposed by Zeshan. A pilot study conducted in 2014 revealed that there were as many instances of morphologically-negated signs accompanied by a negative headshake as instances of headshake-less negative signs (Rutkowski et al., 2015). That clearly indicated that headshaking is optional when negation is conveyed manually and suggested that headshakes reinforce negation in negative contexts rather than grammatically mark it. This observation was endorsed when we analyzed the whole annotated dataset – out of all occurrences of morphologically-negated signs (4060 cases) 47% were accompanied by a headshake, while 53% occurred without this non-manual feature. This observation raises the question of the point at which corpus researchers are likely to experience plateau effects for various linguistic phenomena, making adding new annotations redundant. This has implications both for the planning of other research on the same language and corpus, and for proposed research using other corpora of other sign languages.

In the process of annotation of the PJM data for the purposes of the present study we inserted more than 18,000 tags into the NMNS_HEAD tier and the same amount into the NEG_MAN tier in the iLex software. While the whole dataset contained 244,000 tokens, we note that negation concerns approximately 7% of all produced signs only. Among the 18,000 NMNS_HEAD tags, we identified more than 15,000 instances of left-to-right head movements and approximately 3,000 negative manual signs without any kind of head movements. Out of all instances of headshakes (approximately 10,000), 73% were classified as negative (this is the count for movements appearing in clauses as well as loosely in the discourse). This count is bigger than for the Australian data, probably because of a broader dataset, but we still find a lot of examples of headshakes with other functions (27%, i.e., approximately 2700 cases). What is interesting, in the whole dataset we found only 450 examples of manual verbs negated solely by means of a headshake. However, there were as many as 1900 cases of headshakes accompanying morphologically-negated verbs. This provides further support for the claim that the PJM headshake's nature might be gestural rather than grammatical.

As for nodding, it was only marked in the second round of annotation, in the data subset. It occurred 6 times in negated clauses and was articulated together with a morphologically negated sign (NOT*KNOW, NOT*WAS, NOT*HAVE or NOT*PERMITTED, meaning something like: 'yes, it is forbidden'), once it was co-articulated with the manual sign meaning 'NO' and 3 times occurred having a phatic function and marking the interlocutor's acceptance of the signer's negative utterance. Since in the subset of 140 clips we identified 10 cases of nodding in the negative contexts, we might expect approximately 50 such cases in the whole dataset. We also found 3 cases of head tilting in the negative contexts, but no instances of head turning.

After the second round of annotation (functional tagging), we found out that when a headshake does not target the co-articulated manual sign, it most frequently plays the role of a response marker (330 out of nearly 900 tagged cases). It is also fairly common for the headshake to be a meta-comment of the built narration (180 cases) or to serve a purely phatic function (113 cases – most of them being

articulated without any manual sign), which is not surprising, given the conversational character of the corpus material. Negative imperatives appeared only 4 times.
In the light of the presented observations, we can no longer certainly state whether it is possible to classify Auslan and PJM accordingly to Zeshan's typology (the fact that negative constructions in sign languages exhibit much more variation than could be predicted on the basis of Zeshan's typology is also noted by Oomen and Pfau, 2017, and Huddlestone, 2017). We tentatively suggest that headshaking appears not to have been incorporated into the linguistic systems of Auslan and PJM in any unexpected way, serving rather gestural than grammatical function in the discourse. The analyzed corpus data suggests that headshaking behavior in negative environments may not be all that different from the way in which the hearing people in their vocal communication use it.

4. Some Observation on Preparing and Annotating Sign Language Corpus Data

Besides shedding some new light on negation in sign languages, our work on the reported project also resulted in a number of observations that are relevant when it comes to building and using sign language corpora. The most important ones are listed below:

1. Our experience shows that the process of annotation benefits greatly if both signers' videos are visible to the annotator at the same time whenever possible. This is due to the conversational character of the data. The interlocutor's feedback is important for understanding the discourse context of most signed utterances. This is only a matter of settings in annotation programs, but is often overlooked by researchers, which can lead to some disorientation while glossing and translating signed texts.
2. It is advisable to create a separate annotation tier for each phenomenon under inspection. In the PJM project, head movements relevant from the point of view of the negation study were marked independently from other non-manuals that may have been co-articulated with negative sentences. This helped to avoid confusion, as some head movements did not play any role in expressing negation.
3. While studying negation we learned the importance of not only tagging for relevant head movements, but also paying attention to the overall syntactic structure. Dividing the data into clauses (or clause-like units) is crucial for analyzing what is being negated: the constituent, the clause, or some discourse presupposition. If no annotations are made to the corpus above the level of the individual sign at all (e.g., phrase level, clause level, or sentence level) then it is seriously limited in being able to serve as a basis for linguistic research. This is why the division into clause-like units was included into the annotation process of the PJM Corpus and is now the second step in the annotation workflow.
4. Written translations may be of great help when trying to locate areas of the text that include some device expressing negation (as the relevant translation is likely to involve a negative expression). This method speeds up the process of locating non-manual only expressions of negation. Other than this, one can only search visually for such cases, but they are often easy to miss.
5. It is useful to introduce a ø (zero) tag meaning "this sign/clause was inspected but there is nothing interesting happening here when it comes to the study in question". If this kind of annotation is omitted, annotators and researchers don't know if something has been done in the particular place or not, and it could lead either to enormous waste of time (by requiring the data to be re-inspected) or to significant numbers of missing annotations and hence empty cells if annotations are exported to spreadsheets for processing.
6. It is important to carry out quantitative analyses of the data at several stages of the study in order to control whether the obtained results are changing with new material analyzed. There is no need to add new annotations if the material has hit the plateau effect and is sufficient for providing answers for the research questions posed.

5. Conclusions

Conducting linguistic research on the basis of corpus data definitely adds to our understanding of sign languages. Analyzing extensive datasets might provide new counter-evidence to claims made exclusively on the basis of grammaticality judgments or elicitation. Usage data may not be easily reconciled with intuitions and assumptions about how sign languages function and what their grammatical characteristics are like. The corpus-based study presented in this paper finds more variety in negation patterns than previously described in typological studies. On the other hand, conducting corpus research is time-consuming and, in order to provide credible linguistic data, has to be thoroughly planned and carried out.
We hope that our remarks will be of use to researchers that plan on carrying out detailed analyses of sign language phenomena on the basis of corpus material.

6. Acknowledgements

The PJM corpus project was launched in 2010. Its first phase was supported financially by Poland's National Science Center (*Narodowe Centrum Nauki*) under the project *Iconicity in the grammar and lexicon of Polish Sign Language* (*PJM*) (grant number: 2011/01/M/HS2/03661) and by the Foundation for Polish Science (*Fundacja na rzecz Nauki Polskiej*) under the project *Grammatical categorization through space and movement in Polish Sign Language* (grant number: 1/2009). The second phase is currently financed by the Polish Ministry of Science and Higher Education (*Ministerstwo Nauki i Szkolnictwa Wyższego*) under the National Program for the Development of Humanities (*Narodowy Program Rozwoju Humanistyki* – project title: *Multi-layered linguistic annotation of the corpus of Polish Sign Language* (*PJM*); grant number: 0111/NPRH3/H12/82/2014).
The Polish part of the negation study was supported by a research grant from Poland's National Science Center (project title: *Manual and non-manual markers of negation in Polish Sign Language* (*PJM*), grant number: 2014/13/N/HS2/02872).

7. Bibliographical References

Crasborn, O. & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. 6th International Conference on Language Resources and Evaluation* (*LREC'08*). Paris: ELRA, pp. 39–43.

Crasborn, O. & Zwitserlood, I. (2008). The Corpus NGT: an online corpus for professionals and laymen. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. 6th International Conference on Language Resources and Evaluation* (*LREC'08*). Paris: ELRA, pp. 44–49.

Hanke, T. & Storz, J. (2008). iLex – A database tool for integrating sign language corpus linguistics and sign language lexicography. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. 6th International Conference on Language Resources and Evaluation* (*LREC'08*). Paris: ELRA, pp. 64–67.

Hanke, T., König, L., Wagner, S. & Matthes, S. (2010). DGS Corpus & Dicta-Sign: The Hamburg Studio Setup. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. 7th International Conference on Language Resources and Evaluation* (*LREC'10*). Paris: ELRA, pp. 106–109

Huddlestone, K. (2017). A preliminary look at negative constructions in South African Sign Language: Question-Answer clauses. *Stellenbosch Papers in Linguistics*, 48 (A. Andrason, T. Biberauer (Eds.), *Iets kleins/'n Klein ietsie: A festschrift to honour Johan Oosthuizen on the occasion of his retirement*), pp. 93–104.

Filipczak, J., Johnston, T., Kuder, A., Mostowski, P. & Rutkowski, P. (2015). Between non-manual gesture and grammar: Is headshake a negation marker in Polish Sign Language (PJM) and Australian Sign Language (Auslan)? Paper presented at the Workshop: Nonmanuals at the Gesture Sign Interface (NaGSI), Goettingen, Germany.

Johnston, T. (2010). From archive to corpus: transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15(1), pp. 97–114.

Oomen, M. & Pfau, R. (2017). Signing NOT (or not): A corpus-based study on negation in Sign Language of the Netherlands. *Linguistic Typology*, 21(1), pp. 1–51.

Pfau, R. (2015). The grammaticalization of headshakes: From head movement to negative head. In A.D.M. Smith, G. Trousdale, R. Walereit (Eds.), *New directions in grammaticalization research*. Amsterdam: Benjamins, pp. 9–50.

Rutkowski, P., Kuder, A., Filipczak, J. & Mostowski, P. (2015). Analiza suprasegmentalnych wykładników negacji w polskim języku migowym (PJM) jako przykład wykorzystania metod korpusowych w badaniach nad komunikacją wizualno-przestrzenną [An analysis of suprasegmental markers of negation in Polish Sign Language (PJM) as an example of using corpus methods in research on visual-spatial communication]. In P. Stalmaszczyk (Ed.), *Metodologie językoznawstwa. Od dialektologii do dialektyki* (*Łódzkie Studia z Językoznawstwa Angielskiego i Ogólnego / Łódź Studies in English and General Linguistics*). Łódź: Wydawnictwo Uniwersytetu Łódzkiego, pp. 109–130.

Rutkowski, P., Kuder, A., Filipczak, J., Mostowski, P., Łacheta, J. & Łozińska, S. (2017). The design and compilation of the Polish Sign Language (PJM) Corpus, In P. Rutkowski (Ed.), *Different faces of sign language research*. Warsaw: University of Warsaw, Faculty of Polish Studies, pp. 125–151.

Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S. & Cormier, K. (2013). Building the British Sign Language Corpus. *Language Documentation & Conservation*, 7, pp. 136–154.

Zeshan, U. (2004). Hand, head, and face: Negative constructions in sign languages. *Linguistic Typology*, 8(1), pp. 1–58.

Zeshan, U. (Ed.) (2006). *Interrogative and negative constructions in sign languages*. Nijmegen: Ishara Press.

Queries and Views in iLex to Support Corpus-based Lexicographic Work on German Sign Language (DGS)

Gabriele Langer, Anke Müller, Sabrina Wähl
Institute of German Sign Language and Communication of the Deaf, University of Hamburg
Gorch-Fock-Wall 7, 20354 Hamburg, Germany
E-mail: {gabriele.langer, anke.mueller, sabrina.waehl}@uni-hamburg.de

Abstract

In the DGS-Korpus project the corpus is being used as the basis for lexicographic descriptions of signs in dictionary entries. In this process the lexicographers start from the data and type entry structures as found in the annotation database. While preparing a dictionary entry much of the work consists of manually going through a number of single tokens viewing the original data and available annotations. Findings are then categorised and summarised. However, a number of decisions and descriptions are also supported by pre-defined searches and views on the data. Supported areas include lexicographic lemmatisation (lemma sign establishment), selection of citation forms and variants, grammatical behaviour of signs, collocational patterns of use, regional distribution patterns and distribution of lexical or formational variants over different age groups. While we are still in the process of exploring the possibilities of a sign language corpus for lexicography, searches and views that have proven useful for our work are exemplified in this paper with regard to dictionary entries.

Keywords: corpus-based lexicography, corpus searches, German Sign Language (DGS)

1. Introduction

One of the central aims of the DGS-Korpus project is the compilation of a corpus-based dictionary of German Sign Language (DGS). The basis for a lexicographic description of signs is the reference corpus that was collected within the project for lexicographic and other purposes (Blanck et al., 2010). Corpus data is accessed through the annotational and lexical database and working environment iLex (Hanke & Storz, 2008). During the lexicographic work on preparing an entry it is essential that the available data can be viewed easily and quickly. While working on different entries similar basic analytical questions regarding a sign's properties re-occur with regard to different signs. It is helpful that such questions can be answered quickly through pre-defined queries and views on the data (cf. Atkins & Rundell, 2008: 103, 104).

2. Corpus-based Dictionary of DGS

The dictionary of DGS being produced within the DGS-Korpus project is the first corpus-based dictionary of DGS. Its aim is a description of signs and their use as found in the corpus. Lexicographic descriptions and decisions are informed directly by the analysis of the available corpus data. Dictionary entries include example sentences for the described sign senses directly taken from the originally recorded corpus material.

3. DGS-Korpus Data

The corpus of the DGS-Korpus project was intended to serve as basis for the dictionary from the very beginning. Some elicitation tasks (*Subject Areas, Calender Task, Regional Specialities* and *Elicitation of Isolated Signs*) were specifically included for this purpose (cf. Nishio et al., 2010). The data consists of signed conversations, narrations, discussions, retellings, and other sign uses of 330 informants filmed between 2010 and 2012. Informants from all over Germany were included and balanced for gender, four age groups and 13 regions. For the balancing of regions the estimated population size of sign users was taken into account. Informants were filmed in pairs in one-day sessions. Nearly 560 hours of signing were recorded, up to now 64 hours are completed for basic lemmatisation and annotation. Lemmatisation and annotation is ongoing.[1] Material that is not yet or will not be lemmatized is to a large degree at least translated and can be searched via the translations for specific concepts. In some cases this leads to spot annotations of relevant passages. The corpus size is now approx. 465.000 tokens (23.02.2018).

4. iLex

iLex is the annotational and lexical database and working environment that is used in the DGS-Korpus project for annotating corpus data. It is – up to now – also the only tool that we use to access and view the DGS-Korpus data for the purpose of a lexicographic description of signs.[2]
In iLex, type entries are created to represent abstract sign types to which occurrences of signs (i.e. tokens) are linked. Two type entries can be related to each other in superordinate-subordinate relationship: each type can have only one superordinate type, while a superordinate type may have a number of subordinate types. The user of iLex can define the number of type levels they need in order to set up their data structures.
iLex also provides the user with the possibility to define, store, and re-use SQL-queries and to generate

[1] Lemmatisation here is token-type-matching and an important part of the basic annotation. Lemmatisation in the lexicographic sense may follow different criteria to decide on which elements are attributed lemma sign status and receive their own dictionary entry. To avoid confusion, lemmatisation in the lexicographic sense will here be called *lemma sign establishment* following a suggestion of Svensén (2009: 94).

[2] Other ways to access the data are described and discussed in Jahn et al. (2018) also in this issue.

distributional maps and other visualisations directly from the data (Hanke, 2016).[3]

5. Annotational Type Structures

It is helpful to know how the corpus data is structured in our annotational database in order to better understand the views shown in this paper. In the DGS-Korpus project we use two main and two secondary type levels to build a hierarchical *type structure*[4] that pre-structures the token evidence belonging to one sign.

A sign – an abstract independent meaningful unit of DGS – with all its forms and meanings is represented by a type entry at the highest level, called *sign* in the iLex type structure. A sign is defined and distinguished from other signs by its abstract form, overall range of meaning, and by its underlying image, in cases where its form has been iconically motivated. A *sign* type entry is represented in iLex by a unique gloss[5] and a specific citation form noted in HamNoSys. Instantiations (i.e. tokens) of a sign usually can be identified as belonging to a specific conventional use or meaning of this sign. Such established uses of a sign are modelled in our iLex database as subtypes called *lexemes*. *Lexemes* are subtype entries that are subordinate to *sign* entries. They group tokens that share one of the conventional meanings of the sign. *Lexeme* entries are specified by a unique gloss, a HamNoSys noting their citation form, and a rough indication of their conventional meaning through the assignment of concepts. Each *lexeme* belongs to exactly one *sign* while one (polysemous) *sign* can have a number of *lexemes* attached to it. Tokens that belong to the sign but cannot or have not yet been identified as established uses are not matched at the *lexeme* level but on the *sign* level within the type structure.[6]

The two main type levels of *signs* and *lexemes* are used in the basic annotational lemmatisation process, the token-type-matching. In a second step we also want to gain an impression of the different realisations of the form a sign can take – be it formational (or phonological) variation, grammatical or iconic modification or simply the range of realisations due to performance factors. For that purpose tokens differing from the citation form of *signs* or *lexemes* are grouped by adding recurring form features to the *sign* or *lexeme* gloss. These features name the difference to the citation form by the way of descriptive categories with feature values that are added to the *sign* or *lexeme* gloss. The categories are called *qualifiers* and the resulting groupings are called *qualified signs* or *qualified lexemes*. In iLex, these groupings are modelled as type entries subordinate to *signs* or *lexemes.* Their form is described by HamNoSys notation.[7] Tokens connected to a *qualified sign* or *qualified lexeme* are instantiations of the corresponding superordinate type or subtype.

In the views *sign* glosses are marked by an additional -$SAM at the end to indicate glosses of the highest type level (e.g. TIME1-$SAM). Lexical variants and non-related signs that share the same gloss word are distinguished by numbers (e.g. OR1 vs. OR2). Formational variants are distinguished from each other by letters following the number (e.g. OLD2A vs. OLD2B). *Sign, lexeme*, *qualified sign,* or *qualified lexeme* type entries are created in the lexical database only when needed for annotation. Thus, the hierarchical type structure belonging to one *sign* and the pre-sorting of tokens through that hierarchy provide a first structured view on the corpus data for the respective *sign* (cf. fig. 2 and fig. 8 in apx.).

6. Preparing dictionary entries

A dictionary entry aims at describing the typical uses of words – or in our case signs – disregarding rather untypical uses in order to inform the addressee of the dictionary about how to understand or use a respective item (e.g. Atkins & Rundell 2008: 54, 272). To this aim the lexicographer interprets, weighs and summarizes corpus findings and sometimes other sources of information in a user-oriented, standardised way.

Preparing a corpus-based dictionary entry involves a number of different steps. Atkins and Rundell (2008: 98-103) describe the first stage of this process as the *analysis* stage. The lexicographer reviews and analyses the available data and stores all noticeable facts about the sign in a pre-dictionary database which will serve at a later stage – the *synthesis* stage – as the basis for writing the actual dictionary entry.

In the DGS-Korpus project we are now at the analysis stage of preparing entries based on corpus data.[8] For this, it is essential that all data concerning a sign can be

[3] Self-written queries can be located at different spots within iLex. For example *display filters* define the information to be displayed in lists of items such as type lists; *lists* define the contents to be seen in tabs of display windows, e.g. a subtype list for a supertype or a token list in the type window.

[4] Terms we use to refer to entities and elements in our iLex database are indicated by italics.

[5] iLex uses a relational database. Token-type-matching is internally done via automatically generated IDs. Therefore glosses do not need to bear the function of IDs (as the ID-glosses in Johnston, 2008). They are nothing but unique labels for sign types for the practical handling while working with the data. A gloss can be easily changed without any effect on the lemmatisation and results. In the way we set up our structures in iLex, a constraint prohibits that two different types can be given the same gloss. The new gloss will appear in all transcripts, type entries and views automatically. For the purpose of making the DGS-Korpus publicly accessible for an international audience, types have also been given English glosses. In this paper, we have changed most views to display English glosses instead of German ones.

[6] Each token belonging to a *lexeme* at the same time also belongs to the superordinate *sign* and thus can be identified by both glosses – the *lexeme* gloss or the *sign* gloss, depending on which level of abstraction one wants to focus. This is what we call *double glossing* (cf. Konrad et al., 2012).

[7] As annotation is an ongoing process, *qualifiers* have been defined and introduced to iLex for a number of recurring form features corresponding to modification and variation kinds, but not for all occurring ones.

[8] This paper focuses on corpus data. We also use data obtained by an online survey system on signs and their use called the DGS-Feedback. For how we use data from the DGS-Feedback see Wähl et al. (2018) in this issue.

accessed easily from the corpus. We store our findings in a FileMaker database which at the moment serves as our pre-dictionary database. This database usually contains more information on a sign than what will appear in an actual dictionary entry. Elements of the proto-entry in this pre-dictionary database are marked for publication. Preliminary entries are then produced from exports of this database converted by scripts into an html structure. Representative studio recordings of single signs and original corpus examples prepared in iLex for publication are added to the preliminary entries.
In the remainder of the paper, we will discuss different queries and views in iLex that we have created and found helpful for analysis and decision-making when working on dictionaries entries. We will do so by roughly following the different steps of the workflow. Examples are given to show how corpus data can help answer questions that are relevant to lexicographic decisions and descriptions of signs. Our topic is not to discuss how to construct SQL-queries but what kind of views and pre-stored queries have proven useful in the process.

6.1 Lemma Sign Selection

Sign types are taken as lemma sign candidates and frequency counts help to estimate which types have enough data to enter the lexicographic process. Figure 1 shows a *filter* displaying a type list with a frequency count of attached tokens and the number of subtypes (*lexemes*) with a token count of 25 and above. We consider 25 tokens a minimum number necessary for a description of sign senses of a conventional use of a sign (i.e. a *lexeme*).[9]

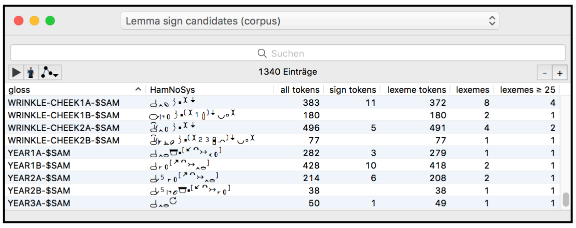


Figure 1: Part of lemma sign candidate list

6.2 Establishment of Lemma Signs

While establishing a lemma sign many different aspects have to be considered. The starting point for an entry is the corpus evidence as it presents itself in the pre-structured way of the annotational database. First, the lexicographer needs to decide, according to the lemmatisation rules of the dictionary, what portion of the data is best described together in one entry or where to split the data into more than one entry. Lexicographic decisions can follow different rules than annotational decisions and may result in a partly different grouping of evidence (cf. Langer et al., 2016). Lemma sign establishment requires an overview of the type structure of the lemma sign candidate and possibly related signs (variants similar signs). The list view of the lemma sign candidates (see fig. 1) already gives an impression of related signs. The type structure of these signs can be displayed and compared within list views showing the *lexemes* and their *qualified forms* (see fig. 2).
The *signs* a) WRINKLE-CHEEK1A-$SAM and b) WRINKLE-CHEEK1B-$SAM are formationally and iconically related. With respect to those characteristics, they might be phonological variants and constitute one single lemma sign. The signs both show *lexemes* with the meaning[10] 'old' but only sign a) can also mean e.g. 'woman', 'mother' or 'grandma'. Additionally, only the *lexeme* of sign b) with the meaning 'old' can undergo numeral incorporation[11]. So, difference in evidenced meanings and grammatical behaviour are two good reasons to describe the signs as two different lemma signs and thus in two entries. Nevertheless cross-references between the dictionary entries will be made because of their iconic and formational relationship. Thus, dictionary users can easily find similar and related signs.

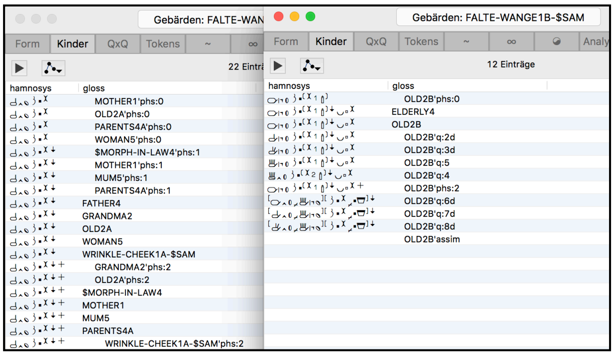


Figure 2: (*Qualified*) *lexemes* of the *signs* a) on the left and b) on the right

6.3 Main Variant and Citation Form

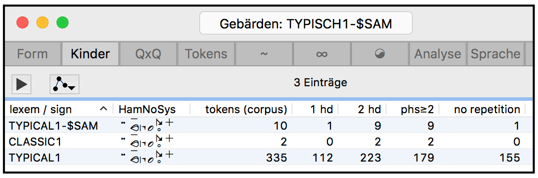
The data to be described in one particular dictionary entry may contain several different sign forms – be it form variants, morphologically relevant modifications or just differences due to performance. For example phonetic variance such as one-handed vs. two-handed occurrences or non-morphological variance in movement repetition can be observed. In a dictionary entry, one form is chosen to represent the whole lemma sign in all its occurring forms. This form is called *lemma* or *citation form*. The lexicographer needs to decide which variants to display in the entry, which variant to choose as main variant and which form of this variant to choose as citation form. Summarised listings of occurring sign forms with token counts are available for a description of form variants in the dictionary (see fig. 3).
Criteria for the choice of main variant can be a higher frequency, broader regional distribution, and broader range of meaning. The corpus data can help to decide what the main variant might be. A query sorting out the

[9] This is a somewhat arbitrary number we chose relating to Sinclair (2005: 11) who suggests a minimum of at least 20 instances necessary for an outline description of the behaviour of a not particularly ambiguous word. Depending on the properties of the respective *lexeme* more evidence might be necessary.

[10] At this stage Word Sense Disambiguation (WSD) still has to be conducted. Thus the given meanings are preliminary and do not specify the whole range of senses the signs may have in the dictionary entry.

[11] This morphological difference is marked by the *qualifier q:* and the corresponding number being incorporated (see fig. 2). The letter *d* behind the number signifies that the handshape includes the thumb.

frequency of forms can be executed, meaning-related as well as form-related. The *sign* TYPICAL1-$SAM is phonologically simple and exhibits some phonetic variation with respect to handedness and repetition. Figure 3 shows an overall distribution of these features and gives an impression of the most frequent forms.



lexem / sign	HamNoSys	tokens (corpus)	1 hd	2 hd	phs≥2	no repetition
TYPICAL1-$SAM		10	1	9	9	1
CLASSIC1		2	0	2	2	0
TYPICAL1		335	112	223	179	155

Figure 3: Summary for number of hands and repetition

The main variant of the lemma sign TYPICAL1-$SAM seems to be two-handed. As for repetition, the picture is not straight forward; but taking into account the many contextual or performative uses of one-handed forms in general, the main variant tends to include repetition. Having summaries of evidence for occurring form variations helps the lexicographer to make an informed decision on a citation form for the entry.

6.4 Description of Meaning (WSD)

The core task for the lexicographer is a documentation of the evidenced range of meaning. This is described by the way of dictionary *senses*. Usually this entails looking through corpus data by the way of a KWIC[12] view on the data – that is a selection of concordance lines (Atkins & Rundell, 2008: 311). The lexicographer groups the contextual meanings of the tokens and describes them as senses and sub-senses in the pre-dictionary database (a process also called Word Sense Disambiguation (WSD), cf. Atkins & Rundell 2008: 269).

In iLex, a number of different views on the data are available when working on analysing, categorising, and summarising the meaning range of a sign. While preparing a sign entry, many tokens are reviewed one by one in context. The analyser views both original recording as well as the corresponding annotations. When the token numbers do not allow the analysis for all tokens in detail, the most promising ones are selected. A token list displaying *lexeme* and *qualified lexeme* glosses, mouthings, translations, left and right neighbours, informants, region, and data collection tasks supports making an informed choice – covering a variety of people, regions, subjects and linguistic context (see fig. 4, apx.).

In iLex the view corresponding to a KWIC list is called *tokens in context*. This view is implemented in iLex and can be filled as needed by the iLex user through suitable queries. For selected tokens, it provides important information such as a sign string, mouthings, translation. Additional information on informant, region and elicitation task is displayed in the lower part for the activated line (see fig. 5, apx.).

Since the DGS data are not written in nature, they do not allow for quick browsing. Available annotations can support but not fully replace viewing the original movie. The original recording corresponding to the selected *tokens-in-context* line can be opened and viewed quickly.

Another view we find helpful for the WSD is the view of frequent left and right neighbours that is described in the next chapter. Collocational patterns can help to identify different uses of a sign with regard to meaning (cf. Atkins & Rundell 2008: 301-304; Kilgarriff, 2012: 7).

6.5 Collocational Patterns

Co-occurrence patterns not only help to distinguish sign senses, but they are also used to identify collocational patterns, idiomatic phrases, and compounds or compound-like combinations. This is supported by a view listing frequent left and right neighbours of the *lexemes* of a given *sign* type (see figure 6, apx.).[13] The view lists neighbours of a lexeme when this combination appears at least five times in the corpus. The view also shows the mutual information score and the number of pattern tokens and informants for that pattern.

Figure 6 (see apx.) shows the co-occurrence results for the type structure of TIME1-$SAM (ZEIT1-$SAM). Marked in blue are combinations that can be interpreted as compound-like sign strings shadowing the elements of German compounds usually accompanying the signs in form of respective mouthings, e.g. YEAR TIME1 (German compound *Jahres|zeit*) or TIME1 PRESSURE (German compound *Zeit|druck*). Often, these compound-like patterns are not fixed combinations of two particular lemma signs but more dynamic combinations. For example, there are three different lexemes YEAR1, YEAR2 and YEAR3 (with in total six formational variants) contributing to the pattern YEAR TIME1.

Marked in red are combinations with number signs or number-incorporating signs used for indicating the time of the day. Marked in orange are other combinations also relating to the time of the day. Two central meanings of the sign can be identified through the green combinations 'good time' and 'beautiful time' versus the yellow combinations NONE/MORE/MUCH-OR-MANY TIME1, TO-NEED TIME1, TIME1 BARELY, and TIME1 FOR. The green combinations are typical of the sense that can be described as 'a specific period in history or a person's life'. The yellow combinations are typical for the following sense of 'time': 'a resource that is needed or available to conduct some activity and that can be plentiful, limited, scarce or lacking.' Lists of frequent neighbours can also indicate typical arguments or argument groups of predicate signs and possibly it will also be helpful to detect idiomatic phrasal structures. In lexicography, we use the list of frequent neighbours

[12] KWIC = keyword in context.

[13] This view on the corpus data is based on a formula that has been used in the Sketch Engine up to September 2006 to determine the mutual information score (MI) (see Lexical Computing, 2015: 2). In our annotational database sometimes even small formational differences are differentiated by new type or subtype entries in order to be able to detect and analyse regional differences. One result of this is smaller token numbers for each grouping. For the purpose of WSD, such finer-grained distinctions are conflated into more general groupings in the co-occurrence list. This is done in a rather coarse way by leaving off additional numbers and letters behind the gloss names (which normally indicate lexical and formational variants) when running the co-occurrence analysis (cf. fig. 6, apx.). The analysis is sensitive with regard to meaning and therefore is run on the *lexeme* level. Individual neighbouring subtypes contributing to the pattern are listed in the last column of the view.

especially for WSD. Dictionary entries are planned to include typical collocational patterns and compound-like combinations, which are selected from the neighbours' list.

6.6 Grammatical Behaviour

In the annotation process, some of the annotated *qualifiers* refer to or serve as a possible indicator of grammatical properties of a sign, such as numeral incorporation or spatial behaviour. There are list views that serve to find candidates for grammatical behaviour and show the existence or presumed non-existence of selected *qualifier* features. One view summarises all tokens with one particular feature regardless of the individual values. *Qualifiers* directly signalling grammatical behaviour are for example *source/goal* or *goal*, expressing form features of indicating or so-called agreement verbs (see fig. 7 for the *sign* TO-VISIT-OR-TO-ATTEND1-$SAM); or *body location*, referring to the morphological change of place of articulation with respect to body parts, or *qualifiers* noting number incorporation (cf. fig. 2). Indicators of possibly grammatical behaviour are *qualifiers* for *phases* (i.e. repetition) as they can point to aspect forms, as well as alterations of speed and size.

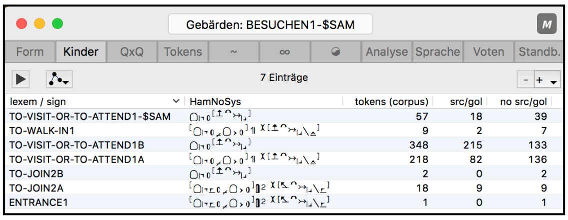


Figure 7: Summary for feature *source/goal*

For the interpretation of the figures it is important to consider the stage of the annotation process (adding *qualifiers* is part of the detailed annotation at a later stage). This means that not all modified forms (*src/gol*) may already be marked. Also, as the form from centre to front is the citation form in the annotational database, only those tokens receive a *qualifier* whose forms differ from the citation form.

For a closer look, other list views show the total range of token forms of a given type and its subtypes, giving the count of tokens for each individual form (see fig. 8, apx.). This view is helpful to gain an overview of the sign's possible characteristics (and how they are distributed with respect to different senses as roughly represented by subtypes), but also to pick out interesting cases. This way usage restrictions or subsenses connected to certain formational behaviour can be found. Since annotation and lemma revision are ongoing processes, the presented situation is not yet fully fledged but helps to detect grammatical behaviour and larger formational sign classes.

6.7 Regional Distribution

A dictionary description of the regional distribution of a lemma sign is easily supported by the rendering of maps as visual representations of distributional patterns (Hanke et al., 2017). For the lexicographic work we regularly use two kinds of maps that show either token numbers or numbers of informants using the sign(s), that is types or subtypes, in question. The grading of regionality follows our data collection subregions within Germany. Maps can easily be rendered directly from the data in iLex by marking the respective types or subtypes in a type list and selecting the desired pre-stored map kind.

The first map kind visualizes the use of the selected type or subtype by indicating the number of tokens (or, if desired, informants) by a colouring from white to yellow to orange to dark red in eight steps. This map gives a good impression on where the sign or lexeme is used and where the core areas of use are. See for example the number of informants using the *lexeme* OR3 in figure 9 (see apx.).

The second kind of maps visualises and contrasts the use of a cluster of lexical and formational variants for presumably the same concept. For each subregion the number of informants using the types or subtypes (or if selected: number of tokens used in that region) are displayed as a pie chart. The pie charts' size is relative to the total number of items and regions are coloured with the colour of the item with the strongest evidence. See for examples the variant cluster for the *lexemes* with the meaning 'or' (fig. 10, apx.).

The map kind 2 (cluster of *lexemes*) shows regional differences and confirms that the regional distribution as shown in the map kind 1 is not the result of still missing data from other regions but truly a result of the use of different variants.

6.8 Age Related Sign Use (Language Change)

Language change is another aspect to be considered while writing an entry, as information on age groups and their preference of signs or sign variants is valuable information on sign use. Signs that show less and less usage along age groups descending from "senior" to "junior" may be prone to vanish and therefore are marked in the entry as "dated". This can occur with respect to specific meanings of a sign (as represented by *lexemes*), or to all meanings. In the latter case the whole sign would be regarded as dated. To detect patterns of language change, clusters of *lexemes* of the same meaning can be compared with respect to the four age groups established.[14] It is advisable to look at clusters and not only isolated lexemes, to minimise effects of chance distribution and get more reliable results (cf. Hanke et al., 2017). For example, the *lexemes* TO-MOVE2 and TO-MOVE1 from different *sign* types are both used to denote 'to move (change of residence)'. The two signs differ in handshape and show a considerable age effect, which we can see via doughnut charts that visualise age distribution with possible clusters. The count can either be on tokens or on different informants. Informant count is more significant here. Two different views have proven helpful. Fig. 11 shows the distribution of informants from the four age groups per *lexeme*. A balanced overall distribution of informants on age groups (with respect to the signs compared) is a prerequisite for a reliable result, which is met in this example (see fig. 11, apx., doughnut on the right).

[14] Based on a date of reference (01.01.2011) the corpus informants were grouped into age groups. The years of birth of the defined age groups lie between 1981-94 for the defined age group 18-30, 1966-1980 for the age group 31-45, 1965-1951 for the age group 46-60 and ≤ 1950 for the age group 61+. People from the cohort ≥1995 have not been included in the corpus because they were not of age at the time of recording.

Another type of doughnut chart view highlights the number of informants of a certain age group using TO-MOVE2 or TO-MOVE1 (see fig. 12 in apx.). Here, the increase of use of TO-MOVE1 can be seen from left to right, where the left doughnut represents the oldest informants and the right one the youngest. With those instruments to analyse the use of signs with respect to age groups, possible trends can be discovered and documented.

7. Conclusion

Corpus-based lexicography of a sign language is a comparatively new field as larger corpora of these non-written languages are now becoming available. Not all of the tools and methods developed for written languages can be directly or effortlessly applied to sign corpora. However, even today, corpus data can already answer many questions on sign use more reliably than it was possible before. The process of developing and experimenting with useful ways to annotate, analyse, summarise and visualise sign corpus data for the needs of sign lexicography is ongoing, and we continuously improve and add to our queries and views on the data. From our experience, we are convinced that in the future sign lexicography will benefit even more from corpora when annotation conventions and analysis methods are further developed.

8. Acknowledgements

This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

9. Bibliographical References

Atkins, B.T.S., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography.* Oxford: Oxford University Press.

Blanck, D., Hanke, T., Hofmann, I., Hong, S.-E., Jeziorski, O., Kleyboldt, T., König, L., König, S., Konrad, R., Langer, G., Nishio, R., Rathmann, C., Vorwerk, S., Wagner, S. (2010). The DGS Corpus Project. Development of a Corpus Based Electronic Dictionary German Sign Language – German. Poster presented at the Theoretical Issues in Sign Language Research Conference (TISLR 10), Sept 30 - Oct 2, 2010 at Purdue University, Indiana, USA.

Hanke, T., Konrad, R., Langer, G., Müller, A., Wähl, S. (2017). Detecting Regional and Age Variation in a Growing Corpus of DGS. Poster presented at the 9th International Corpus Linguistics Conference (CL 2017) pre-conference workshop "Corpus-based Approaches to Sign Language Linguistics: Into the Second Decade", Jul 24, 2017 at Birmingham, UK.

Hanke, T. (2016). Towards a Visual Sign Language Corpus Linguistics. In E. Efthimiou, E. Fotinea, T. Hanke, J. Hochgesang, J. Kristoffersen, & J. Mesch (Eds.), *Corpus Mining. Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages. 10th International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia.* Paris: ELRA, pp. 89--92.

Hanke, T., Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, & E. Thoutenhoofd (Eds.), *Construction and Exploitation of Sign Language Corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages. 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco*. Paris: ELRA, pp. 64--67.

Jahn, E., Konrad, R., Langer, G., Wagner, S., Hanke, T. (2018). DGS-Korpus Project: Different Formats for Different Needs. In this issue.

Johnston, T. (2008). Corpus linguistics and signed languages: No lemmata, no corpus. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, & E. Thoutenhoofd (Eds.), *Construction and Exploitation of Sign Language Corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages. 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco*. Paris: ELRA, pp. 82--87. Paris: European Language Resources Association.

Kilgariff, A. (2012). Using Corpora [and the Web] as Data Sources for Dictionaries. Online resource; URL: https://www.sketchengine.co.uk/wp-content/uploads/Using_corpora_2012.pdf [Accessed 2018-02-19].

Konrad, R., Hanke, T., König, S., Langer, G., Matthes, S., Nishio, R., Regen, A. (2012). From form to function. A database approach to handle lexicon building and spotting token forms in sign languages. *Interaction between Corpus and Lexicon. Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages. 8th International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey*. Paris: ELRA, pp. 87--94.

Langer, G., Troelsgård, T., Kristoffersen, J., Konrad, R., Hanke, T., König, S. (2016). Designing a Lexical Database for a Combined Use of Corpus Annotation and Dictionary Editing. In E. Efthimiou, E. Fotinea, T. Hanke, J. Hochgesang, J. Kristoffersen, & J. Mesch (Eds.), *Corpus Mining. Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages. 10th International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia.* Paris: ELRA, pp. 143--152.

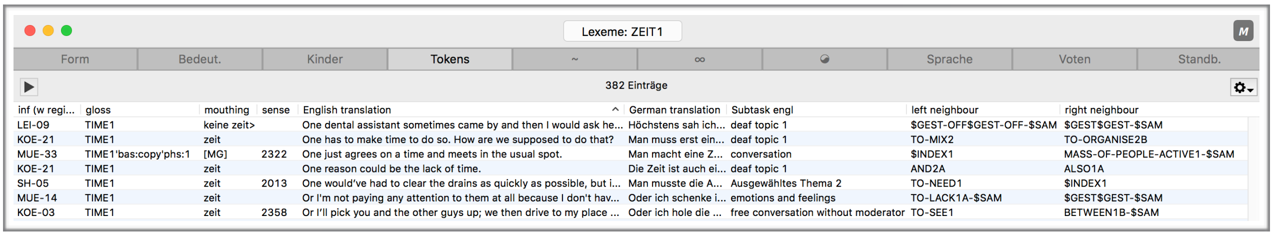
Lexical Computing Ltd. (2015): Statistics used in the Sketch Engine. July 8, 2015. Online resource; URL: http://www.sketchengine.co.uk/wp-content/uploads/ske-statistics.pdf [Accessed 2017-01-18].

Sinclair, J. (2005). Corpus and Text – Basic Principles. In M. Wynne (Ed.), *Developing Linguistic Corpora: a Guide to Good Practice.* Oxford: Oxford Books, 1--16. Available online from: http://ota.ox.ac.uk/documents/creating/dlc/ [Accessed 2018-02-19].

Svensén, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making.* Cambridge: Cambridge University Press.

Wähl, S, Langer, G., Müller, A. (2018). Hand in Hand – Using Data from an Online Survey System to Support Lexicographic Work. In this issue.

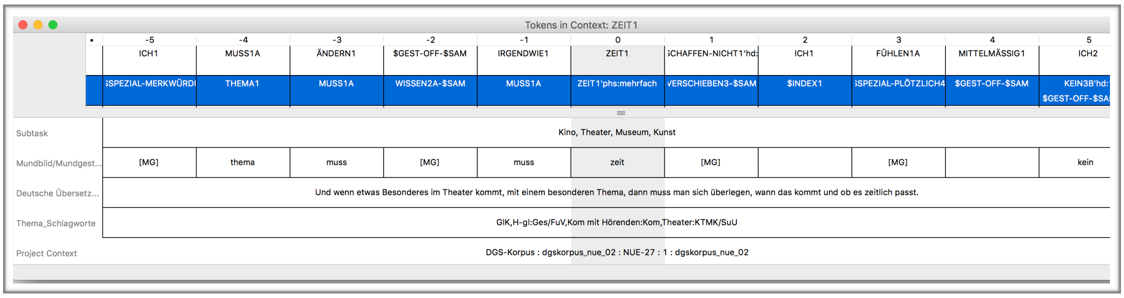
10. Appendix



Lexeme: ZEIT1 — Form | Bedeut. | Kinder | Tokens | ~ | ∞ | ◕ | Sprache | Voten | Standb. — 382 Einträge

inf (w regi...	gloss	mouthing	sense	English translation	German translation	Subtask engl	left neighbour	right neighbour
LEI-09	TIME1	keine zeit>		One dental assistant sometimes came by and then I would ask he...	Höchstens sah ich...	deaf topic 1	$GEST-OFF$GEST-OFF-$SAM	$GEST$GEST-$SAM
KOE-21	TIME1	zeit		One has to make time to do so. How are we supposed to do that?	Man muss erst ein...	deaf topic 1	TO-MIX2	TO-ORGANISE2B
MUE-33	TIME1'bas:copy'phs:1	[MG]	2322	One just agrees on a time and meets in the usual spot.	Man macht eine Z...	conversation	$INDEX1	MASS-OF-PEOPLE-ACTIVE1-$SAM
KOE-21	TIME1	zeit		One reason could be the lack of time.	Die Zeit ist auch ei...	deaf topic 1	AND2A	ALSO1A
SH-05	TIME1	zeit	2013	One would've had to clear the drains as quickly as possible, but i...	Man musste die A...	Ausgewähltes Thema 2	TO-NEED1	$INDEX1
MUE-14	TIME1	zeit		Or I'm not paying any attention to them at all because I don't hav...	Oder ich schenke i...	emotions and feelings	TO-LACK1A-$SAM	$GEST$GEST-$SAM
KOE-03	TIME1	zeit	2358	Or I'll pick you and the other guys up; we then drive to my place ...	Oder ich hole die ...	free conversation without moderator	TO-SEE1	BETWEEN1B-$SAM

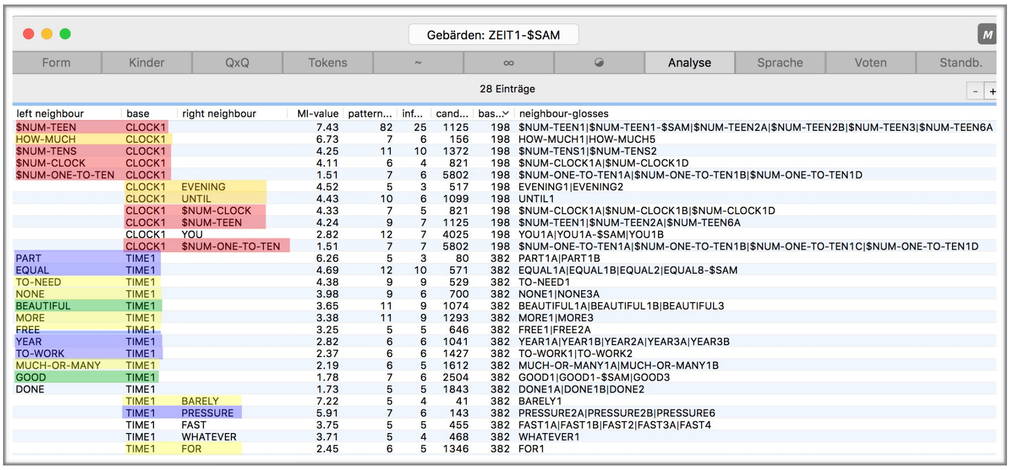
Figure 4: Token list of the *lexeme* TIME1 with relevant information for WSD



Tokens in Context: ZEIT1

	-5	-4	-3	-2	-1	0	1	2	3	4	5
	ICH1	MUSS1A	ÄNDERN1	$GEST-OFF-$SAM	IRGENDWIE1	ZEIT1	;CHAFFEN-NICHT1'hd:	ICH1	FÜHLEN1A	MITTELMÄSSIG1	ICH2
	;SPEZIAL-MERKWÜRDI	THEMA1	MUSS1A	WISSEN2A-$SAM	MUSS1A	ZEIT1'phs:mehrfach	VERSCHIEBEN3-$SAM	$INDEX1	;SPEZIAL-PLÖTZLICH4	$GEST-OFF-$SAM	KEIN3B'hd: $GEST-OFF-$SA
Subtask	Kino, Theater, Museum, Kunst										
Mundbild/Mundgest...	[MG]	thema	muss	[MG]	muss	zeit	[MG]		[MG]		kein
Deutsche Übersetz...	Und wenn etwas Besonderes im Theater kommt, mit einem besonderen Thema, dann muss man sich überlegen, wann das kommt und ob es zeitlich passt.										
Thema_Schlagworte	GIK,H-gl:Ges/FuV,Kom mit Hörenden:Kom,Theater:KTMK/SuU										
Project Context	DGS-Korpus : dgskorpus_nue_02 : NUE-27 : 1 : dgskorpus_nue_02										

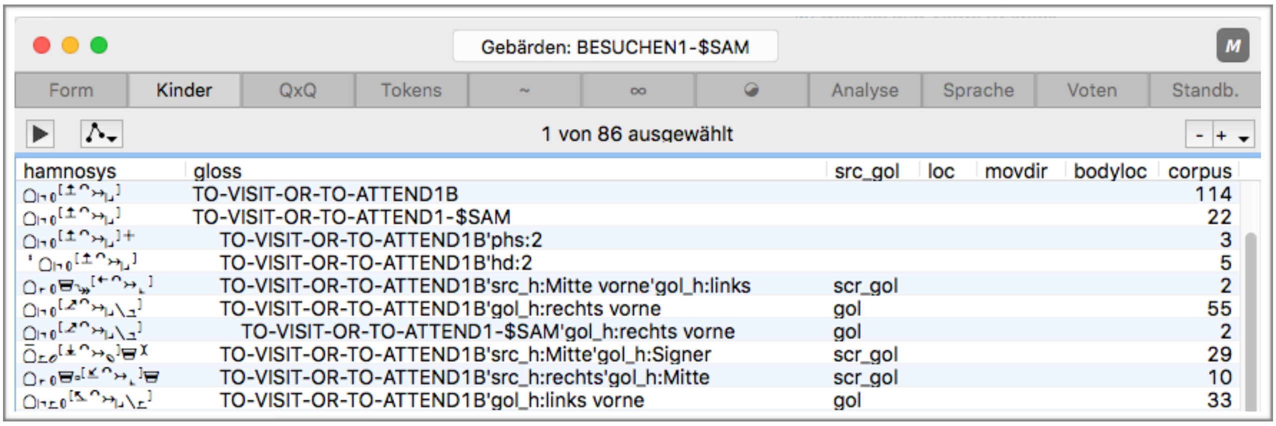
Figure 5 : View *tokens in context*



Gebärden: ZEIT1-$SAM — Form | Kinder | QxQ | Tokens | ~ | ∞ | ◕ | Analyse | Sprache | Voten | Standb. — 28 Einträge

left neighbour	base	right neighbour	MI-value	pattern...	inf...	cand...	bas...	neighbour-glosses
$NUM-TEEN	CLOCK1		7.43	82	25	1125	198	$NUM-TEEN1\|$NUM-TEEN1-$SAM\|$NUM-TEEN2A|$NUM-TEEN2B\|$NUM-TEEN3|$NUM-TEEN6A
HOW-MUCH	CLOCK1		6.73	7	6	156	198	HOW-MUCH1|HOW-MUCH5
$NUM-TENS	CLOCK1		4.25	11	10	1372	198	$NUM-TENS1\|$NUM-TENS2
$NUM-CLOCK	CLOCK1		4.11	6	4	821	198	$NUM-CLOCK1A\|$NUM-CLOCK1D
$NUM-ONE-TO-TEN	CLOCK1		1.51	7	6	5802	198	$NUM-ONE-TO-TEN1A\|$NUM-ONE-TO-TEN1B|$NUM-ONE-TO-TEN1D
	CLOCK1	EVENING	4.52	5	3	517	198	EVENING1|EVENING2
	CLOCK1	UNTIL	4.43	10	6	1099	198	UNTIL1
	CLOCK1	$NUM-CLOCK	4.33	7	5	821	198	$NUM-CLOCK1A\|$NUM-CLOCK1B|$NUM-CLOCK1D
	CLOCK1	$NUM-TEEN	4.24	9	7	1125	198	$NUM-TEEN1\|$NUM-TEEN2A|$NUM-TEEN6A
	CLOCK1	YOU	2.82	12	7	4025	198	YOU1A|YOU1A-$SAM|YOU1B
	CLOCK1	$NUM-ONE-TO-TEN	1.51	7	7	5802	198	$NUM-ONE-TO-TEN1A\|$NUM-ONE-TO-TEN1B|$NUM-ONE-TO-TEN1C\|$NUM-ONE-TO-TEN1D
PART	TIME1		6.26	5	3	80	382	PART1A|PART1B
EQUAL	TIME1		4.69	12	10	571	382	EQUAL1A|EQUAL1B|EQUAL2|EQUAL8-$SAM
TO-NEED	TIME1		4.38	9	9	529	382	TO-NEED1
NONE	TIME1		3.98	9	6	700	382	NONE1|NONE3A
BEAUTIFUL	TIME1		3.65	11	9	1074	382	BEAUTIFUL1A|BEAUTIFUL1B|BEAUTIFUL3
MORE	TIME1		3.38	11	9	1293	382	MORE1|MORE3
FREE	TIME1		3.25	5	5	646	382	FREE1|FREE2A
YEAR	TIME1		2.82	6	6	1041	382	YEAR1A|YEAR1B|YEAR2A|YEAR3A|YEAR3B
TO-WORK	TIME1		2.37	6	6	1427	382	TO-WORK1|TO-WORK2
MUCH-OR-MANY	TIME1		2.19	6	5	1612	382	MUCH-OR-MANY1A|MUCH-OR-MANY1B
GOOD	TIME1		1.78	7	6	2504	382	GOOD1|GOOD1-$SAM|GOOD3
DONE	TIME1		1.73	5	5	1843	382	DONE1A|DONE1B|DONE2
	TIME1	BARELY	7.22	5	4	41	382	BARELY1
	TIME1	PRESSURE	5.91	7	6	143	382	PRESSURE2A|PRESSURE2B|PRESSURE6
	TIME1	FAST	3.75	5	5	455	382	FAST1A|FAST1B|FAST2|FAST3A|FAST4
	TIME1	WHATEVER	3.71	5	4	468	382	WHATEVER1
	TIME1	FOR	2.45	6	5	1346	382	FOR1

Figure 6: Frequent left and right neighbours of the *sign* TIME1-$SAM



Gebärden: BESUCHEN1-$SAM — Form | Kinder | QxQ | Tokens | ~ | ∞ | ◕ | Analyse | Sprache | Voten | Standb. — 1 von 86 ausgewählt

hamnosys	gloss	src_gol	loc	movdir	bodyloc	corpus
[illegible]	TO-VISIT-OR-TO-ATTEND1B					114
[illegible]	TO-VISIT-OR-TO-ATTEND1-$SAM					22
[illegible]	TO-VISIT-OR-TO-ATTEND1B'phs:2					3
[illegible]	TO-VISIT-OR-TO-ATTEND1B'hd:2					5
[illegible]	TO-VISIT-OR-TO-ATTEND1B'src_h:Mitte vorne'gol_h:links	scr_gol				2
[illegible]	TO-VISIT-OR-TO-ATTEND1B'gol_h:rechts vorne	gol				55
[illegible]	TO-VISIT-OR-TO-ATTEND1-$SAM'gol_h:rechts vorne	gol				2
[illegible]	TO-VISIT-OR-TO-ATTEND1B'src_h:Mitte'gol_h:Signer	scr_gol				29
[illegible]	TO-VISIT-OR-TO-ATTEND1B'src_h:rechts'gol_h:Mitte	scr_gol				10
[illegible]	TO-VISIT-OR-TO-ATTEND1B'gol_h:links vorne	gol				33

Figure 8: Summary of sign forms of TO-VISIT-OR-ATTEND-$SAM with token counts (segment)

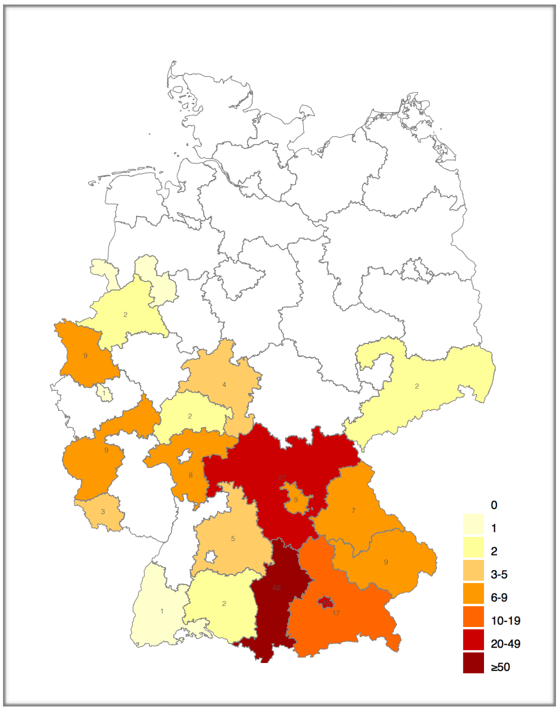


Figure 9: Map (tokens) for *lexeme* OR3

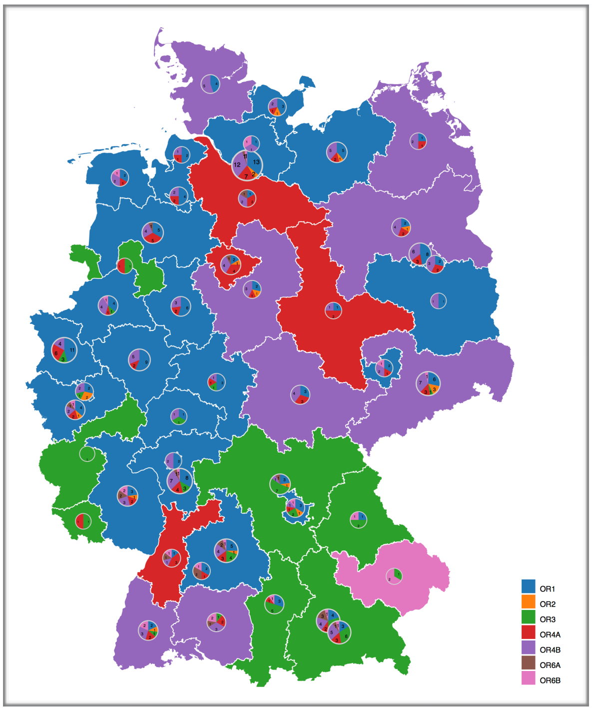


Figure 10: Map (informants) for variant cluster „or“

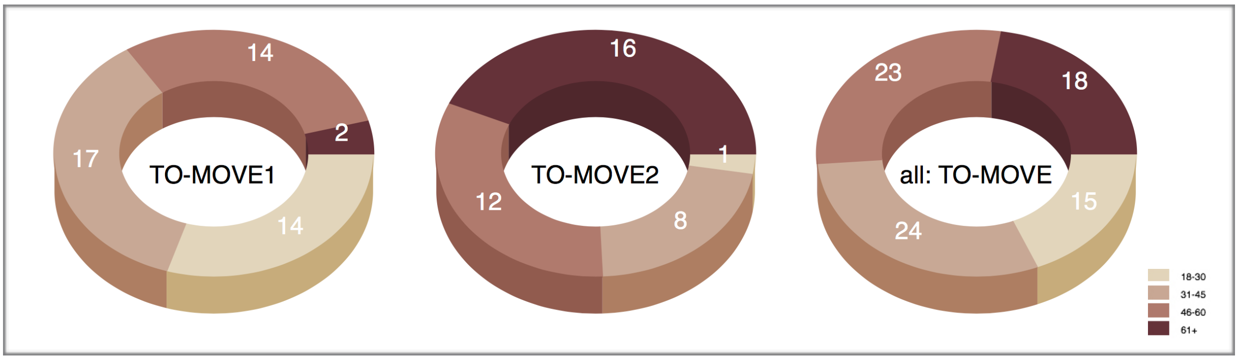


Figure 11: Doughnut charts (informants' age groups per *lexeme*) for „to move“

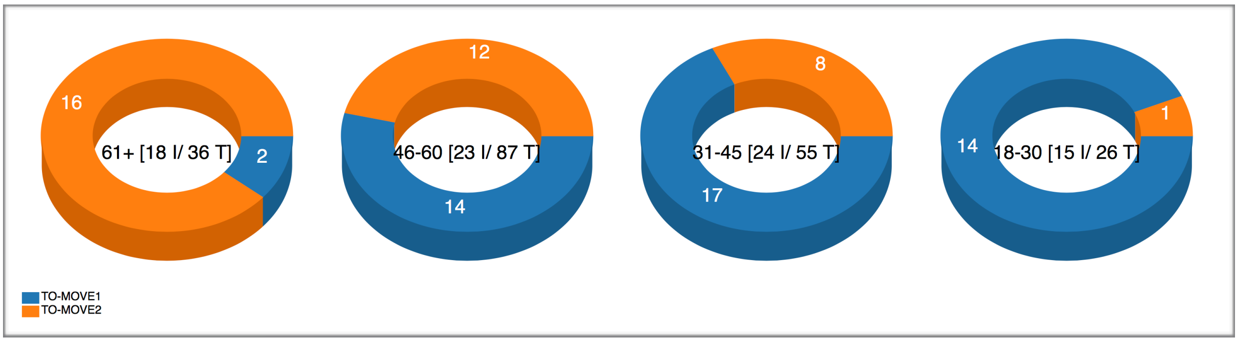


Figure 12: Doughnut charts (*lexemes* per informants' age group)

Which Picture?
A Methodology for the Evaluation of Sign Language Animation Understandability

Vonjiniaina Domohina MALALA[1], Elise PRIGENT[1], Annelies BRAFFORT[1], Bastien BERRET[2]

[1]Limsi, CNRS, Paris Saclay University – [2]CIAMS, Paris-Sud University, Paris Saclay University
[1]LIMSI, Campus d'orsay bat 508, F-91405 Orsay cedex, France
[2]CIAMS, Campus d'orsay bat 335, F-91405 Orsay cedex, France
{elise.prigent, annelies.braffort}@limsi.fr, bastien.berret@u-psud.fr

Abstract

The goal of our study is to explore which information is essential to understand virtual signing. To that aim, we developed an online test to assess the comprehensibility of four different versions of signers: a baseline version with a real human signer, a most complete version of a virtual signer, and two degraded versions of a virtual signer (one with non-visible hands and one without movements of head/trunk). Each video showed the description of a picture in French Sign Language (LSF). After having seen the video, participants had to find which picture had been described among 9 pictures displayed. The originality of our approach was to include two types of confusable pictures on the response board. One was supposed to induce errors by confounding the lexical signs and the other by confounding the spatial structure of the picture. In this way, we explored the effect of hiding hands and blocking trunk/head on the comprehension of lexicon and spatial structure.

Keywords: Sign Language, SL animation, Virtual signer, Evaluation, Visual Perception

1. Introduction

This paper deals with the evaluation of Sign Language (SL) generation technology. We focused on French SL (LSF) generation based on the animation of a Virtual Signer (VS) using real human movements captured from a motion capture system (Benchiheub et al., 2016).

SL is a visuo-gestural language that uses eyes, face, torso, arms and hands movements to convey meaning. With actual technology, all these movements cannot be replicated accurately and faithfully in virtual signing whereas they are likely necessary for understanding the signing content. This may account for the poor understandability of virtual signers by Deaf people. Therefore, to produce understandable signing, we must determine which visual information is most critical and has to be perfectly animated on the VS. Thus, the questions raised in this study are: What motion information is essential to understand virtual signing? To what extent does the manipulation, simplification or withdrawal of some information affect understanding?

To provide objective answers to these questions, we have to design a method allowing to acquire quantitative measures about visual perception and comprehension of VS (with different qualities) by participants.

In this paper, we introduce the method that was designed and used in a cognitive psychology study related to the visual perception of movements in SL.

The paper is structured as follows: section 2 is dedicated to a review of studies in psychology and computer science about perception and comprehension of human and virtual signers; section 3 details our methodology and the design of the platform used; section 4 proposes a discussion and gives an example of the kind of interesting results that have been acquired thanks to this method.

2. Visual perception of Sign Language

2.1 Perception of Human Signers

Emmorey, Thompson & Colvin (2009) have shown that both Deaf native signers (Deaf people with SL as native language) and hearing beginning signers (who had completed between 9 and 15 months of SL instruction) look at the observed signer's face more than 80% of the time. But these authors also showed that beginning signers move more frequently their attention from the face to the hands than native signers. Native signers would focus their attention on the eyes while retaining the ability to integrate the information from the manual parameters with peripheral vision (Morford et al., 2008). Despite native Deaf signers focus their attention on the face, they recognize more quickly the signs conveyed by the hands than beginning signers (Morford & Carlson, 2011). This confirms that this is rather the peripheral vision of signers that is used to perceive the rapid movements of the hands and fingers, while central vision is used to perceive movements located on the face.

Thus, according to Muir (2005), a good spatial resolution of the image at the face level (with good temporal resolution maintained throughout the video) is necessary for understanding SL videos. For this author, it may be possible to reduce the quality of the peripheral region, including the body and hands (when away from the face), while retaining the quality of the perceived video.

2.2. Perception of Virtual Signers

The use of Virtual Signers (VS) brings many advantages over videos of real signers. They are anonymous and can be interactive (Kipp et al., 2011b). Nevertheless, their usability is limited by the low level of comprehension by the observers (Kennaway et al., 2007). Most VS are developed by researchers who are not experts of SL linguistics and who tend to create "pleasant" VS, sometimes forgetting that the VS is also a language that convey informa-

tion with movements of the articulators that can be very precise. That is why we must identify which component of the model needs more precision, for optimal understanding (Kipp et al., 2011a).
Moreover, the creation of VS animations remains a difficult task because the movements of the different parts of the body must be well synchronized and it is difficult to reproduce all the spatiotemporal parameters of SL, in particular the non-manual parameters. Criticisms regarding VS have often pointed out these parameters (gaze, facial expression, movement of the mouth, head and bust) (Kipp et al., 2011a). Paradoxically, great attention is usually put on the movements of the hands to facilitate the understanding of SL. For instance, Alexanderson & Beskow (2015) proposed to use a low-cost technology using fewer markers in the animation of the movements of the hands of VS, thereby obtaining a recording of hand movements of less complete/accurate. The results of this study showed that despite the reduction in the amount of information, the comprehensibility and the clarity of the signs was not altered compared to the animation with more markers.
Regarding eye movements, an eye-tracking study showed that when native Deaf people observe human signers, the fixation time of the face is greater than when they observe VS. Accordingly, there is less gaze displacements between the face and the body when observing a human signer rather than a VS (Kacorri et al., 2014).
These previous studies suggest two main results: first, there is a difference about visual exploration (of face and body) of users when observing human or virtual signers, with different levels of quality; second, the comprehension but also the visual exploration used can differ as function of the observer's SL expertise.
In order to determine the parameters of SL that must be modelled more precisely for the optimisation of the VS, we explored observers' comprehension of different types of signers: human, virtual with different qualities by manipulating different relevant parameters. We also explored the impact of the observer's SL expertise on their comprehension. Our VS was animated using motion capture of a human signer and not using synthetic animations, thus guaranteeing data very close to the initial human signing.

3. Methodology

From previous studies, we know that user-based evaluation of SL generation comprehensibility requires many precautions during the design step, regarding the identification of the socio-linguistic profile of the participants and avoid using to much text in order to keep the participants concentrated on SL.
There is no standard process for assessing the comprehensibility of an LSF statement. Generally, simple categories are proposed to evaluate globally the understandability and naturalness, sometimes grammatical correctness, using for example numeral scales or glosses[1] as possible responses given by the participants (Kipp et al., 2011a).
Huenerfauth et al. (2008) have proposed an original process, that consists to use short movies. Each movie gives a dynamic interpretation of an utterance such as "The man walk next to the woman". The participant had to match each SL animation with one movie among three. This approach can provide a more reliable rating of understandability, but it cannot be used for any kind of utterances.

3.1 Our set-up: an online test with complete and degraded animations

As previously mentioned, our objective is to determine which parameters of SL must be modelled more precisely for the optimisation of the VS. To evaluate quantitatively the relevance of different body parts on the SL comprehensibility, a method consists to alter the animations and compare the perception, such as in (Huenerfauth & Lu, 2010) regarding the location of signs, or in (Gibet et al., 2011) for facial expression and gaze. We have used the same kind of method, while trying to add a more reliable way to measure the understandability.
We used Cuxac's model (2000) to determine the relevant parameters. According to this model, we hypothesized that the lack of handshapes should result in more difficulty to identify the lexical signs, while the lack of body and head movements should result in more difficulty to figure out the global structure of the picture, which is described by "showing", in the signing space in front of the signer, the spatial organisation of the picture scene, implying in many cases rotations and movements of the head and the torso. Because many studies focused on lexical signs comprehension, we propose here to explore the comprehension of signs related to more depicting structures, such as size and shape descriptions or localisation of entities in the signing space. Hence, we measured the impact of two degradations of the virtual signers on the comprehension of LS description, and more precisely on the comprehension of lexical signs and depicting signs[2] respectively. In our study, one version of the animation was realised by hiding the hands and the other one by blocking the trunk and head movements on all degrees of freedom.

In order to allow a relatively large number of persons to participate and collect enough data to conduct statistical analyses, we created an online test via a LimeSurvey server (a web application that enables users to develop and publish online surveys, collect responses and export the resulting data).
We asked participants to watch 8 videos containing picture descriptions in LSF (see section 3.1.). After each video, the participant had to choose the picture described among a set of 9 pictures (see, section 3.2.). This online test was send to Deaf Signer, Hearing Signer and Hearing Non Signer using mailing lists or social networks in France.

[1] Word or set of word expressing the same concept (or at least the closest), i.e. the gloss SCIENCE for the lexical sign representing the concept of *science.*

[2] These types of signs are often referred to as 'classifier' signs. See (Liddell, 2003) for a detailed definition of depicting sign.

3.2 Stimuli

For the creation of visual stimuli, we used the LSF corpus called MOCAP1 (Benchiheub et al., 2015), which in particular contains videos and motion capture data of the description of pictures. An expert made an annotation of the corpus, segmenting the gestural units in the videos, then identifying the lexical signs and the depicting structures, especially those showing size and shape of objects, localisation of objects or spatial relations between objects. Based on these annotations, we chose the 4 descriptions with approximately the same number of lexical and depicting signs in order to create our stimuli (these 4 pictures are illustrated in Figure 1). In this corpus, in addition to a camera, 3D recordings of the movements were made using a motion capture system (Optitrack). These recordings were then used to animate the VS.

The physical appearance of the VS, the color of his skin, the clothes and the background of the video were chosen to have as much resemblance as possible to the original video. From the 3D recordings, a Deaf computer graphist created 3 different versions of VS (Figure 2):

- A complete animation without modification (Complete VS, Figure 2.b).
- A degraded animation with hands hidden by spheres (Handless VS, Figure 2.c)
- A degraded animation by freezing the trunk and head movements (Blocked VS, Figure 2.d)

Since the human signer had no markers on the fingers, the computer graphist manually animated fingers and facial expressions by using the rotoscopy method. So these 3 versions of VS presented facial expressions based on those displayed by the real signer. Thus, for each of the 4 pictures (Figure 1), we obtained 4 videos of the description corresponding to 4 types of signers: human signer, complete VS, handless VS, and blocked VS (Figure 2).

3.3 Modality of response

After each video, a response board with 9 pictures was displayed, and the participant had to choose the picture described in the video. For each video description, 8 confusable pictures were carefully chosen according to the expert's annotations mentioned previously. 4 pictures presented similarities in the lexicon (related to the objects present in the scene), 4 others in the global structure of the described picture.

More precisely, the 9 pictures were composed of:

- 1 picture corresponding to the correct response, the one described in the video (Figure 3: n°6),
- 4 confusable pictures with similar spatial structure (Figure 3: n° 4, 5, 7, 8),
- 4 confusable pictures with similar lexical elements (Figure 3: n° 1, 2, 3, 9).

The same response board was displayed in the two conditions "real signer" and "virtual signer", but with a different ordering of the pictures in the response board.

Because we displayed two degraded versions of VS, we could test whether the handless VS induces more confusion for pictures with similar lexical elements and, conversely, whether the blocked VS induces more confusion for pictures with similar structures.

Figure 1. Example of pictures used to elicite descriptions in the MOCAP1 LSF corpus. These are the 4 ones used in our study.

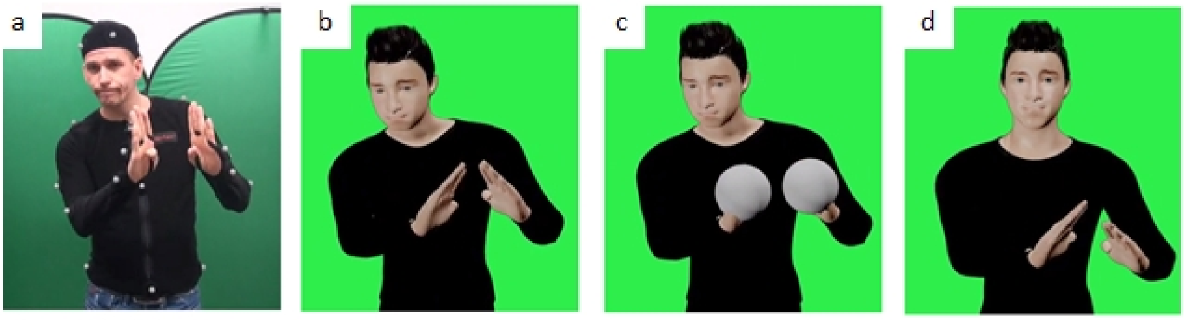


Figure 2. Extracts of videos in the 4 conditions: a) Real signer; b Complete VS; c) Handless VS; d) Blocked VS.

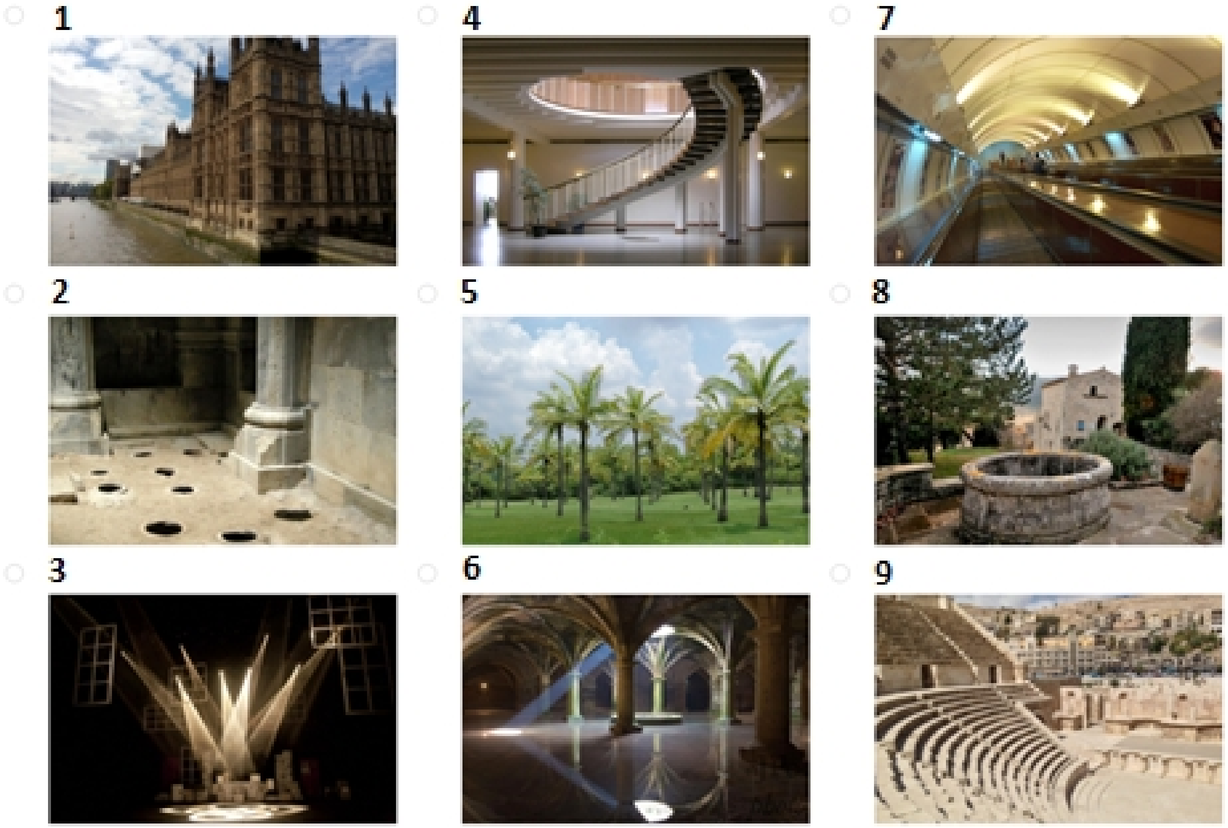


Figure 3. Example of a response board. The picture 6 is the right answer. Pictures 1, 2, 3, 9 are supposed to induce lexical errors. Picture 1 could induce the lexical sign “water”; 2: “hole”; 3: “light”; 9: “Roman”. Pictures 4, 5, 7, 8 are supposed to induce structural errors. Picture 4 could induce a description that illustrates the shape of a hole in the ceiling; 5: shape of pillars and ceiling; 7: shape of vaulted ceiling; 8: shape of the well edge.

3.4 Procedure

We asked participants to run the test on desktop computer or laptop to ensure good viewing conditions of the videos, because the screens of mobile phones or touch pads are too small.

The test lasted about 10 minutes. The first page of the test provided instructions and explanations on the process, and the informed consent of the participants. The test was performed anonymously. Instructions were presented in written form and with a video translation in LSF to facilitate accessibility and understanding of the task by Deaf persons who might present reading difficulties.

Prior to the comprehension test, participants were asked to indicate their age, gender, nationality, hearing status and expertise in LSF. That constituted 3 groups: Deaf Signer (DS), Hearing Signer (HS), Hearing Non Signer (HNS). There were no Deaf Non Signer participants. Depending on the group to which they belong, the participants were directed toward different questions. For example, a DS or HS participant had to answer questions related to his/her level in LSF (according to the European Common European Framework of Reference for Language). A DS

participant had to answer questions related to their place and age of learning of LSF, etc.
The comprehension test was composed of two blocks, the first with 4 videos description of VS and the second with 4 videos description of human signer. Within each block, the order of the 4 videos was randomized. We created 3 different versions of the VS (complete, handless and blocked). So, each participant was randomly oriented to one of the 3 versions of VS (complete, handless or blocked) in block 1 and all participants watched the same block 2 of human signers. Each video lasted between 20 and 25 seconds. For each trial, one video of description of picture in LSF was displayed. Participants could only view the video once. The "play", "stop" and "progress bar" commands were deactivated and backtracking on the web page was not allowed. To prepare the participant to the video, a 4-seconds countdown was displayed before his beginning. Once the video is finished, the participant clicked on the “next” button to access the page containing the response board with 9 pictures (1 good response, 4 “structural” confusable pictures and 4 “lexical” confusable pictures). The participant had to click on the picture that he thinks correspond to the description in the previous video, and then click on the "next" button to move on to the next video. Before starting the comprehension test, a familiarization trial, not included in statistical analyses, was displayed, composed of a video of the same person describing a different image than the ones used in the test.
At the end of the test, the participants had the opportunity to get the number of correct answers they got and to give their impressions on the test and on the VS by using a text field.

4. Discussion

The results of this specific study is not the focus of this paper. But to say a word, the very first analysis gives some insights about the role of movement in understandability. For example, even with the degraded versions of the VS animations and even in the group of hearing non signers participants, some of them were able to find the good response. On the other hand, even Deaf and hearing signers could be disturbed by the degraded versions of the VS but not necessarily in the same manner.
The discussion here is more on the design of the test. The originality of our method is to propose a link between the information degraded in the stimuli (here, hands hidden and blocking trunk and head movements) and the confusable pictures displayed in the response board. This design allows us to measure the impact of the degradation of a visual information on the comprehension of the message, and more precisely on the comprehension of two types of signs, lexical and depicting, by analysing errors made by the participants. The results may provide interesting conclusions both for linguistic and computer science domains. First, they could serve linguistic models by providing information about the relative importance of the movement of specific body parts (face, hand and bust) for the various type of sign (lexical or depicting). Second, this study may provide some new guidelines for the animation of VS. Because synthetic animation of VS does not allow to accurately replicate all the movements of a human signer, a simplification is necessary. So, this kind of study can propose recommendations about simplification of one motion parameter rather than another as a function of the message produced (e.g. lexical signs or depicting signs) and of the expertise in SL of the participants.
Moreover, we have yet some inputs on the way the test could be improved. Actually, near 200 participants have completed the online test. We had much more participants, but responses from non-French participants and those who did not perform the test until the end were excluded from the analyses. Thus, the design of an online test allows to get a sufficient number of participants as well as to perform robust and reliable statistical analysis. However, we have not a balanced size of participants in the 3 groups (Deaf signers, Hearing signers and Hearing non signers). There were less Deaf participants. It also appeared that several participants had only a smartphone and thus were rejected from the test for which we asked to use a desktop computer or a laptop. Also, a limitation is that the participants had the opportunity to give their impressions on the test only by text. That could be a brake for Deaf people who present writing difficulties. We plan to add the possibility to post impressions via a video in follow-up studies of this type.
Another difficulty is related to the duration of the descriptions. They lasted between 20 and 25 seconds, which may seem short, but they contain an important number of elements (between 17 and 26 depicting and lexical signs). Overall, the descriptions are already quite complex. Therefore, there may be some memorisation issues that are part of the difficulty of the task. We assume that this effect, which is the same for all the participants, has no influence on the results and interpretations. However, this is perhaps one of the reasons why some participants did not complete the test. It would be interesting in the future to think about a more playful way of presenting the test, like a serious game for example.

5. References

Alexanderson, S., & Beskow, J. (2015). Towards Fully Automated Motion Capture of Signs – Development and Evaluation of a Key Word Signing Avatar. *ACM Transactions on Accessible Computing*, 7(2), 1-17.

Benchiheub, M.-E., Berret, B., & Braffort, A. Collecting and Analysing a Motion-Capture Corpus of French Sign Language, 7th International Conference on Language Resources and Evaluation - Workshop on the Representation and Processing of Sign Languages (LREC-WRPSL 2016), Portoroz, Slovenia, 7-12.

Cuxac, C. (2000). La Langue des Signes Française (LSF) : les voies de l'iconicité, Paris-Gap, Ophrys, Bibliothèque de *Faits de Langues* n°15 -16, 2000

Emmorey, K., Thompson, R., & Colvin, R. (2009). Eye Gaze During Comprehension of American Sign Language by Native and Beginning Signers. *Journal of Deaf Studies and Deaf Education*, 14(2), 237-243.

Gibet, S., Courty, N., Duarte, K. & Naour, T. (2011), The SignCom System for Data-driven Animation of

Interactive Virtual Signers: Methodology and Evaluation, *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 1:1, 6:1--6:23.

Huenerfauth, M., Zhao, L., Gu, E., Allbeck, J.: Evaluating american sign language generation by native ASL signers. *ACM Transactions on Access Computing* 1(1), 1–27 (May 2008)

Huenerfauth, M. & and Lu, P. (2010), Modeling and synthesizing spatially inflected verbs for American sign language animations, Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility (ASSETS'10), Orlando, Florida, USA, 99-106.

Kacorri, H., Harper, A., & Huenerfauth, M. (2014). Measuring the perception of facial expressions in american sign language animations with eye tracking. In International Conference on Universal Access in Human-Computer Interaction (p. 553–563). Springer.

Kennaway, J. R., Glauert, J. R. W., & Zwitserlood, I. (2007). Providing signed content on the Internet by synthesized animation. *ACM Transactions on Computer-Human Interaction*, 14(3).

Kipp, M., Heloir, A., & Nguyen, Q. (2011a). Sign language avatars: Animation and comprehensibility. In International Workshop on Intelligent Virtual Agents (p. 113–126). Springer.

Kipp, M., Nguyen, Q., Heloir, A., & Matthes, S. (2011b). Assessing the deaf user perspective on sign language avatars, The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility, 107-114.

Liddell, S. K. (2003). Grammar, Gesture, and Meaning in American Sign Language. Cambridge: Cambridge University Press.

Morford, J. P., & Carlson, M. L. (2011). Sign Perception and Recognition in Non-Native Signers of ASL. *Language Learning and Development*, 7(2), 149-168

Morford, J. P., Grieve-Smith, A. B., MacFarlane, J., Staley, J., & Waters, G. (2008). Effects of language experience on the perception of American Sign Language. Cognition, 109(1), 41-53.

Muir, L. J. (2005). Perception of Sign Language and Its Application to Visual Communications for Deaf People. *Journal of Deaf Studies and Deaf Education*, 10(4), 390-401.

Alexanderson, S., & Beskow, J. (2015). Towards Fully Automated Motion Capture of Signs – Development and Evaluation of a Key Word Signing Avatar. *ACM Transactions on Accessible Computing*, 7(2), 1-17.

Benchiheub, M.-E., Berret, B., & Braffort, A. Collecting and Analysing a Motion-Capture Corpus of French Sign Language, 7th International Conference on Language Resources and Evaluation - Workshop on the Representation and Processing of Sign Languages (LREC-WRPSL 2016), Portoroz, Slovenia, 7-12.

Cuxac, C. (2000). La Langue des Signes Française (LSF) : les voies de l'iconicité, Paris-Gap, Ophrys, Bibliothèque de *Faits de Langues* n°15 -16, 2000

Emmorey, K., Thompson, R., & Colvin, R. (2009). Eye Gaze During Comprehension of American Sign Language by Native and Beginning Signers. *Journal of Deaf Studies and Deaf Education*, 14(2), 237-243.

Gibet, S., Courty, N., Duarte, K. & Naour, T. (2011), The SignCom System for Data-driven Animation of Interactive Virtual Signers: Methodology and Evaluation, *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 1:1, 6:1--6:23.

Huenerfauth, M., Zhao, L., Gu, E., Allbeck, J.: Evaluating american sign language generation by native ASL signers. *ACM Transactions on Access Computing* 1(1), 1–27 (May 2008)

Huenerfauth, M. & and Lu, P. (2010), Modeling and synthesizing spatially inflected verbs for American sign language animations, Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility (ASSETS'10), Orlando, Florida, USA, 99-106.

Kacorri, H., Harper, A., & Huenerfauth, M. (2014). Measuring the perception of facial expressions in american sign language animations with eye tracking. In International Conference on Universal Access in Human-Computer Interaction (p. 553–563). Springer.

Kennaway, J. R., Glauert, J. R. W., & Zwitserlood, I. (2007). Providing signed content on the Internet by synthesized animation. *ACM Transactions on Computer-Human Interaction*, 14(3).

Kipp, M., Heloir, A., & Nguyen, Q. (2011a). Sign language avatars: Animation and comprehensibility. In International Workshop on Intelligent Virtual Agents (p. 113–126). Springer.

Kipp, M., Nguyen, Q., Heloir, A., & Matthes, S. (2011b). Assessing the deaf user perspective on sign language avatars, The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility, 107-114.

Liddell, S. K. (2003). Grammar, Gesture, and Meaning in American Sign Language. Cambridge: Cambridge University Press.

Morford, J. P., & Carlson, M. L. (2011). Sign Perception and Recognition in Non-Native Signers of ASL. *Language Learning and Development*, 7(2), 149-168

Morford, J. P., Grieve-Smith, A. B., MacFarlane, J., Staley, J., & Waters, G. (2008). Effects of language experience on the perception of American Sign Language. Cognition, 109(1), 41-53.

Muir, L. J. (2005). Perception of Sign Language and Its Application to Visual Communications for Deaf People. *Journal of Deaf Studies and Deaf Education*, 10(4), 390-401.

6. Language Resource References

MOCAP1 corpus (2015), distributed via Ortolang. perennial identifier https://hdl.handle.net/11403/mocap1

From Design and Collection to Annotation of a Learner Corpus of Sign Language

Johanna Mesch, Krister Schönström
Department of Linguistics
Stockholm University
SE-106 91 Stockholm, Sweden
{johanna.mesch, schonstrom}@ling.su.se

Abstract

This paper aims to present part of the project "From Speech to Sign – learning Swedish Sign Language as a second language" which include a learner corpus that is based on data produced by hearing adult L2 signers. The paper describes the design of corpus building and the collection of data for the Corpus in Swedish Sign Language as a Second Language (SSLC-L2). Another component of ongoing work is the creation of a specialized annotation scheme for SSLC-L2, one that differs somewhat from the annotation work in Swedish Sign Language Corpus (SSLC), where the data is based on performance by L1 signers. Also, we will account for and discuss the methodology used to annotate L2 structures.

Keywords: Learner corpus, annotation, L2 signers, L2 analysis, Swedish Sign Language

1. Introduction

To date, little is known about what learning a sign language, i.e. learning a new language in a new modality, is all about. The creation of a learner corpus of signed language would seem to be an essential step in the right direction in our understanding of the learning process. Such a corpus would have to include a large amount of machine-readable data and be annotated according to guidelines (Granger, Gilquin & Meunier 2015). Learners are used to engaging in classroom activities, i.e. doing role-play with their classmates in order to practice and improve their skills in using the target language, but not to conveying a "genuine" message. A learner corpora can be collected within the context of the university, but it is necessary for its data to be of varying degrees of naturalness, such as simple interviews and the retelling of narratives (Gilquin 2015). Recent research within second language acquisition (SLA) area has pointed to the possibilities of using corpora for research (Wulff 2017). This paper aims to present a learner corpus in Swedish Sign Language that is based on data produced by hearing adult L2 signers, namely the Corpus in Swedish Sign Language as a Second Language (SSLC-L2), which is part of the funded project "From Speech to Sign – learning Swedish Sign Language as a second language" (Schönström & Mesch 2017), and describes ongoing work in specialising the annotation of the SSLC-L2. First, we will present the corpus, including our experiences in developing the corpus. Second, we will account for and discuss the methodology used to annotate L2 structures, i.e. specific L2 structures as well as L2 errors.

2. Corpus Design and Data

2.1. Learner Corpus SSLC-L2

SSLC-L2 is a learner corpus with a longitudinal design for which data from adult second language (L2) learners of SSL has been collected since 2013 (Schönström & Mesch 2017). A parallel corpus for Irish Sign Language and American Sign Language were also established at the same time (Schönström et al. 2015). For the SSLC-L2, the third cohort of learners is being collected, and the last recordings will be completed in Q4 2018. In total, SSLC-L2 will contain data from 38 learners at different stages and times (Table 1). In addition, we have a parallel corpus, i.e. a control group, with nine native signers.

Collection	Recording time	Contact time (total hrs)
Phrase 1	Term 1 September	45
Phrase 2	Term 1 December	125
Phrase 3	Term 2 May	240
Phrase 4	Term 3 December	345

Table 1: Collection of data in phases (recordings and teaching hours)

As part of the collection process, learners are invited to visit our studio individually and to sit with a native signer as interview leader. A learner is asked to reply to some questions and discuss simple issues depending on her/his level, and then to perform retelling tasks (picture and movie task) in four different phrases during a span of 1.5 years. Each session takes 15-20 minutes per person for every phase, and is recorded by the studio's five video cameras.
With the goal of obtaining an authentic data source, we have been striking a balance between free production and elicited tasks in order to broaden possible future investigations of the corpus from a variety of linguistic perspectives. Some of the tasks have been used in the SSLC, providing further opportunities for contrastive comparisons between L1 and L2 signers. The tasks were also given/adjusted according to learners' levels following their developmental points. The *interview* aims to collect conversational/interactional data from the learner, and,

following a longitudinal design, the questions become more complex with time, following the learners' expected linguistic levels, according to the scales of the Common European Framework of Reference for languages (CEFR). *Frog story* consists of selected pictures from the book that aims to elicit basic skills in describing a simple spatial situation. Participants are also given sample pictures from the *transitive utterance elicitation task* of Volterra et al. (1984), with the aim of eliciting orders of elements. *Ferdinand* is a humorous three-picture cartoon strip that aims to elicit narratives in a broad sense. The last one, *The Plank*, is a one-minute sequence from the famous short movie The Plank. This movie is intended to elicit longer narrative sequences at a later stage in the longitudinal collection. For an overview of tasks used in the corpus, see Table 2.

	Month after onset	**Interview**	**Frog, where are you?**	**Transitive utterance**	**Ferdinand**	**The Plank video**
Phase 1	One month	Interview questions A1-A2	Yes	Yes	No	No
Phase 2	Four months	Questions A1-B1	Yes	Yes	No	Yes
Phase 3	Nine months	Questions A1-B2	Yes	No	Yes	Yes
Phase 4	16 months	Questions A1-B2	No	No	Yes	Yes

Table 2: Overview longitudinal data collection and the tasks

2.2. The SSLC-L2 Data

Table 3 shows current data collected so far and the amount of annotated data (id gloss, Swedish translation) (Mesch et al. 2017).

	Edited video data	**Completed annotation files with glosses and translation**
Cohort 1	9:05:58	5:44:02
Cohort 2 (not finished)	6:03:46	
Cohort 3 (not finished)	2:03:24	
	14:53:49	5:44:02

Table 3: Statistics on the annotated SSLC data (as per 20 February 2018)

2.3. Ethical Considerations

The participants are first- and second-year students entering the BA program in sign language and interpreting. Some of them are also beginning students in SSL, having been so for only two terms. They are not doing any assignment for teaching or examination. Participating in the project is voluntary, and only a small portion of each student group is participating. Participants are asked to provide written consent and to complete a background questionnaire (metadata) before participating in the interview and elicitation assignments. The data is sensitive, so it is semi-open only to researchers with permission. A research ethics application has been approved for this project.

3. Annotation Procedures and Outcomes

3.1. Standard for the Annotating of L2 Structures

SSLC-L2 has provided guidelines for annotation (Mesch & Wallin 2015; Wallin & Mesch 2018). These are used in order to maintain annotation standards for ID-glosses in SSLC. All glosses of the SSLC have been annotated with part-of-speech labels (Östling, Börstell & Wallin 2015). The current paper describes some annotation challenges and some aspects of our proposal for additional annotation guidelines that are needed for a specialized L2 corpus. At the first stage, we established an annotation standard for tagging the signs. Here, standard SSLC glosses are selected as target glosses regardless of the produced form, i.e. if they come with phonological or lexical errors, etc.

In the next step, we built a standard for the annotating of L2 structures, including conventions on annotating closely related phenomena, i.e. disfluencies such as silent pauses, fillers (e.g. @hd), unfinished signs (e.g. tree@&) and hesitant pauses (tp@&), etc. Here, we are accounting for annotation solutions related to L2 structures including errors and other disfluencies that appear in spoken languages as well (see, e.g., Gilquin & De Cock 2011). The first L2 structure analysis has been on structures at the lexical and phonological levels. Forthcoming analysis will look at structures on the morphological as well as syntactic level. At this initial stage, we have adopted a *contrastive interlanguage analysis* framework (Granger 2015), that is, we are comparing the L2 output with a parallel group consisting of native SSL signers.

This complex process of annotating L2 structures and errors will be discussed in relation to the existing SLA research area. A special challenge lies in identifying and confirming obligatory contexts for target language structures in sign language mode. In our presentation, we will account for different kinds of manual as well as non-manual L2 structures, including mouth actions, following earlier subcategories (M-type, A-type, etc.), as suggested by Crasborn et al. (2008). Only the B-type has been added to annotate mouth actions functioning as backchannel (lip, laugh, surprised mouth movement) (Wallin & Mesch 2018) because of conversation materials, where the interlocutor uses some mouth actions in order to give backchannel signals to the other signer.

In sum, we created a set of different tiers described in greater detail below (also see Figure 1).

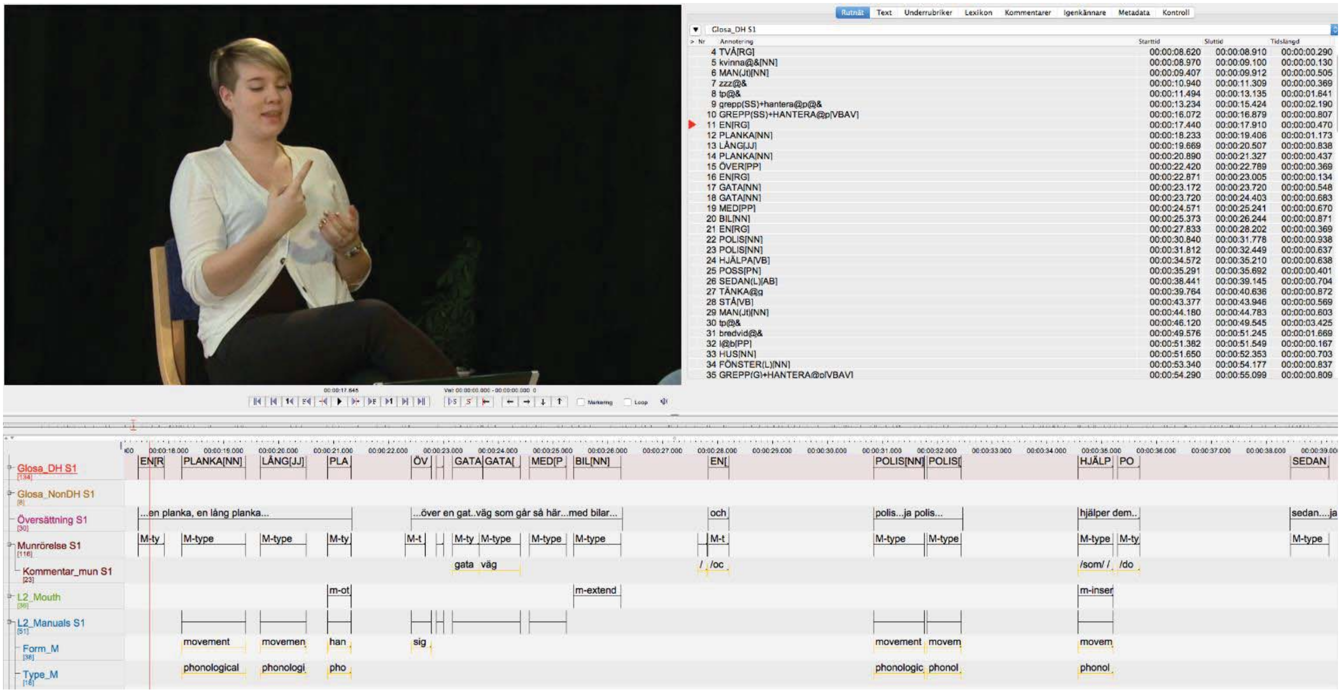


Figure 1: Screen shot of ELAN with the glosses, translation, mouth types and L2 Manual tiers

3.2. Tiers with Manual Information

The basic annotation consists of three tiers for the signer: two for sign glosses with part-of-speech labels, and one for written Swedish translations. One gloss tier is for all signs and other manual utterances (e.g. waving hands, palms up and unfinished signs) with one or two hands. There are also expanded tiers for articulator (one or two hands) and meaning on a ‘child’ tier. Annotating sign glosses is a challenge, as there is partial overlap between the use of gesture and space for meaning and reference, e.g. as the signs in the elicited sequence of the plank movie representing the meaning of ‘carrying the plank’. An L2 signer expresses a sign PLANK ‘plank’ or fingerspells the whole word, but another L2 signer expresses it as depicting sign FORM(SS).DESCRIPTION@p ‘plank’ while using mouthing borrowed from Swedish. When concerning a verb, L2 signers are shown selecting a sign CARRY ‘to carry’ or a description of how to carry a plank, as glossed as a depicting sign GRIP(SS).HANDLE@p.

3.3. Translation Tiers

A tier for translating the content of SSL into Swedish was also established. A hearing native speaker of SSL, a professional sign language interpreter, was hired for the translation work. One challenge has been to mirror some L2 structures and characteristics of particular signing, for example, all the hesitations and thinking pauses, as well as deciphering the signs. We have tried to mirror those structures to some extent, through palm ups, pointings and pause utterances (eeh.., hmm…, etc.).

3.4. Tiers with L2 Analysis

The L2 analysis tiers are divided into two main parts: manual signing and non-manual signing. Annotations of non-manual features for grammatical purposes as well as disfluencies were accounted for in an earlier paper (Schönström & Mesch 2014) and will not be discussed further here. The L2 manual tiers are for the annotation of manual L2 features, including errors and other features typical for L2 signers, see Figure 2. Table 4 shows child tiers for the parent tier L2_Manual in which manual L2 utterances are annotated. This tier focuses on lexical production, including phonological as well as morphological structure and semantic use. Also, a strategy tier was added in order to see which strategies L2 learners use in their sign lexical production. The strategies that have been observed far have been the use of fingerspelling and gestures.

Tier	Tag
Form_M	handshape movement orientation place of articulation sign
Type_M	phonological morphological semantic lexical
Strategy_M	fingerspelling gesture
Comment_M	Free comments

Table 4: Tiers and tags used in the SSLC-L2 for L2 analysis

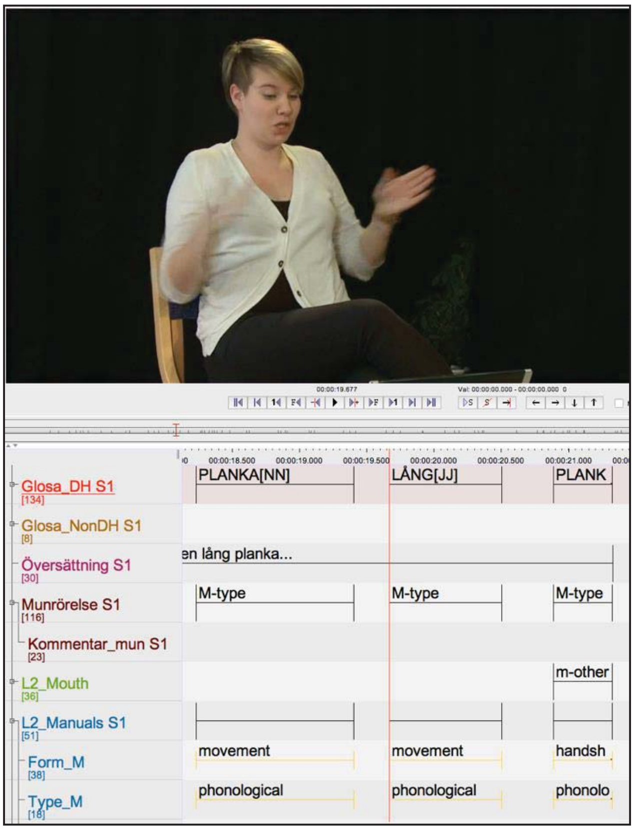


Figure 2: Example with L2 manual tiers

Form_M marks the form of the sign that we have analysed as different from the target language norm. Mostly, this is related to phonological parameters: handshape, movement, orientation, and place of articulation, but also whether the entire lexical sign is erroneous or used in a particular way (for example, if it is related to semantic level).
Type_M defines type of error or derivation in use, i.e. if the form marked in the Form_M tier is related to the phonological, morphological, lexical or semantic level.
While L2_Manual tier focuses on the manual signs for analysing lexical level, there are tiers for analysing syntactic level and mouth actions that are presented in the following sections.

3.4.1. Tiers with L2 Syntactic Analysis

Tier	Description	Tag
L2_Syntatic	Single intransitive argument Transitive Actor Transitive Undergoer Verb Auxiliary verb Non-verbal predicate Obligatory locative complement (Loc)	S A P V{1,2,3} Aux nonV Loc
L2_Clauses	Adverbial Object Relative	

Table 5: Argument tags used in the SSLC-L2 for syntactic analysis

The tier L2_Syntactic tier allows for the annotation of syntactic constructions. Our model is based on Gärdenfors' (2017) work, which is based on the theoretical framework of Role Reference Grammar (Van Valin Jr & La Polla 1997; Börstell et al. 2016), as well as a child tier, L2_clauses, with functional analysis of sub-clause types (relative clauses, object clauses and adverbial clauses marked as in Table 5).

3.4.2. Tiers with L2 Mouth Actions

The category of mouth actions of L2 learners have been annotated on their own tier (Mesch, Schönström, Riemer, & Wallin 2016). Mouth movements borrowed from Swedish (mouthing without sound) are annotated as M-type, and other mouth actions as A/E/4/W/B-types or types for no movement and undefined. There is a very high frequency of M-type, which is a natural "transition" and influence from Swedish for L2 signer. Errors in mouthings appear when an L2 signer tries to describe a sign for a handle verb GRIP(SS).HANDLE@p 'to carry a plank' using M-type, instead of A-type.

4. Analysing the Outcomes of L2 Structures and Errors

Depending on the research agenda and aims, the strength of a corpus-based approach is its sustainability and the possibility of expanding the analysis with new tiers. The base annotation work is time-consuming, but once it's done it is simple to extract statistical information or outcomes of any kind. Here we present some preliminary outcomes for the analysis of errors in manual signing that have been made available through our annotation work (Table 6 and 7).

Form_M (N=91)		
sign	56	34,78%
movement	44	27,33%
handshape	43	26,71%
orientation	8	4,97%
place of articulation	5	3,11%
depicting sign	3	1,86%
phrase	2	1,24%

Table 6: Frequency of form errors or derivations

Type_M (N=91)		
phonological	87	69,60%
morphological	14	11,20%
lexical	12	9,60%
semantic	12	9,60%

Table 7: Frequency of error types

5. Discussion

Our experience creating SSLC-L2 has contributed to new insights. First, regarding the method for data collection. In general, the method for data collection generated a huge amount of data. However, we learned that if one wants to analyse specific constructions, for example, depicting signs, there may be a need to include more elicited tasks specifically aiming for depicting signs in order to elicit more and more varied data. This needs to be taken into consideration in future research. As our data now stands, there are a relatively large amount of depicting signs, but they do not appear in a constant manner, and they are somewhat limited to a relatively small number of "situations".

Second, regarding the annotation of L2 structures, it has been a real challenge, even for us L1 signers, to identify, describe and (if applicable) categorize L2 structures. However, our method of annotation categorisation has helped us to organise the structures. In the future, the L2_Manual tier may need to be separated into more tiers, i.e. in phonological and lexical tiers.

Just like spoken language data, it takes time to establish and code a sign language corpus, and, as we are reaching a critical mass of annotated data, future work will focus on the generation of different research outcomes as well as on producing results.

6. Acknowledgements

This work on the project "Från tal till tecken - att lära sig Svenskt teckenspråk som andraspråk", TATE [From speech to sign – learning Swedish Sign Language as a second language] was supported in part by Riksbankens Jubileumsfond (P16-0371:1). We also acknowledge personnel who have been involved in the technical and annotation aspects of the corpus: Joel Bäckström, Moa Gärdenfors, Ylva Larsson, Nikolaus Riemer Kankkonen and Johan Wallin.

7. Bibliographical References

Börstell, Carl, Mats Wirén, Johanna Mesch & Moa Gärdenfors. 2016. Towards an annotation of syntactic structure in the Swedish Sign Language Corpus. In Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie Hochgesang, Jette Kristoffersen & Johanna Mesch (eds.), *Workshop Proceedings: 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*, 19–24. Paris: ELRA.

Crasborn, O., E. Van Der Kooij, D. Waters, B. Woll & J. Mesch. 2008. Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. *Sign Language and Linguistics (Online)* 11(1). doi:10.1075/sl&l.11.1.04cra.

Gilquin, Gaëtanelle. 2015. From design to collection of learner corpora. In Fanny Meunier, Gaëtanelle Gilquin & Sylviane Granger (eds.), *The Cambridge Handbook of Learner Corpus Research*, 9–34. (Cambridge Handbooks in Language and Linguistics). Cambridge: Cambridge University Press. doi:DOI: 10.1017/CBO9781139649414.002.

Gilquin, Gaëtanelle & Sylvie De Cock. 2011. Errors and disfluencies in spoken corpora: Setting the scene. *International Journal of Corpus Linguistics* 16(2). 141–172. doi:10.1075/ijcl.16.2.01gil.

Granger, Sylviane. 2015. Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research* 1(1). 7–24. doi:10.1075/ijlcr.1.1.01gra.

Granger, Sylviane, Gaëtanelle Gilquin & Fanny Meunier. 2015. *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.

Gärdenfors, Moa. 2017. Syntaktisk struktur i svenskt teckenspråk hos hörande andraspråksinlärare – En analys av ordföljd, bisatser och användning av verb. [The syntactic structure in Swedish Sign Language produced by hearing L2 learners - An analysis of the word order, subordinate clauses and the use of verbs], MA-thesis, Dept. of Linguistics, Stockholm University.

Mesch, Johanna, Krister Schönström, Nicolaus Riemer & Lars Wallin. 2016. The interaction between mouth actions and signs in Swedish Sign Language as an L2. Paper presented at the 12th International Conference on Theoretical Issues in Sign Language Research, TILSR 12, Melbourne, January 6, 2016.

Mesch, Johanna & Lars Wallin. 2015. Gloss annotations in the Swedish Sign Language Corpus. *International Journal of Corpus Linguistics* 20(1). doi:10.1075/ijcl.20.1.05mes.

Schönström, Krister, Matthew Dye, Lorraine Leeson & Johanna Mesch. 2015. Building up L2 Corpora in different signed languages – SSL, ISL and ASL. Poster presented at the 2nd International Conference on Sign Language Acquisition, ICSLA, Amsterdam, The Netherlands, 1-3 July 2015.

Schönström, Krister & Johanna Mesch. 2014. Use of nonmanuals by adult L2 signers in Swedish Sign Language – Annotating the nonmanuals. In Onno Crasborn, Eleni Efthimiou, Evita Fotinea, Thomas Hanke, Julie Hochgesang, Jette Kristoffersen & Johanna Mesch (eds.), *Beyond the Manual Channel. Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages*, 153–156.

Valin Jr, Robert D. Van & Randy J. La Polla. 1997. *Syntax: Structure, Meaning, and Function*. Cambridge: Cambridge University Press.

Volterra, V., A. Laudanna, S. Corazza, F. Natale & E. Radutzky. 1984. Italian Sign Language: The order of elements in the declarative sentence. In F. Loncke, Penny Boyes Braem & Y Lebrun (eds.), *Recent Research on European Sign Languages*. Lisse: Swets & Zeitlinger.

Wallin, Lars & Johanna Mesch. 2018. *Annoteringskonventioner för teckenspråkstexter. Version 6, januari 2018. [Annotation guidelines for*

sign language texts]. Department of Linguistics, Stockholm University.
Wulff, Stefanie. 2017. What learner corpus research can contribute to multilingualism research. *International Journal of Bilingualism* 21(6). 734–753. http://10.0.4.153/1367006915608970.
Östling, Robert, Carl Börstell & Lars Wallin. 2015. Enriching the Swedish Sign Language Corpus with part of speech tags using joint Bayesian word alignment and annotation transfer. In Beáta Megyesi (ed.), *Proceedings of the 20th Nordic Conference on Computational Linguistics (NODALIDA 2015), NEALT Proceedings Series 23*, 263–268. Vilnius: ACL Anthology.

8. Language Resource References

Mesch, Johanna, Krister Schönström, Moa Gärdenfors, Nikolaus Riemer Kankkonen & Ylva Larsson. 2017. Annoterade filer för Korpus i svenskt teckenspråk som andraspråk. Version 1. [Annotation files for the learner corpus of Swedish Sign Language. Version 1.]. Department of Linguistics, Stockholm University.
Schönström, Krister & Johanna Mesch. 2017. Dataset. The project From speech to sign – learning Swedish Sign Language as a second language. Department of Linguistics, Stockholm University.

Scalable ASL Sign Recognition using Model-based Machine Learning and Linguistically Annotated Corpora

Dimitris Metaxas*, Mark Dilsizian*, Carol Neidle**
*Rutgers University Computer Science Department, **Boston University Linguistics Program
*110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, **621 Commonwealth Ave., Boston, MA 02215
dnm@cs.rutgers.edu, mdil@cs.rutgers.edu, carol@bu.edu

Abstract

We report on the high success rates of our new, scalable, computational approach for sign recognition from monocular video, exploiting linguistically annotated ASL datasets with multiple signers. We recognize signs using a hybrid framework combining state-of-the-art learning methods with features based on what is known about the linguistic composition of lexical signs. We model and recognize the sub-components of sign production, with attention to hand shape, orientation, location, motion trajectories, plus non-manual features, and we combine these within a CRF framework. The effect is to make the sign recognition problem robust, scalable, and feasible with relatively smaller datasets than are required for purely data-driven methods. From a 350-sign vocabulary of isolated, citation-form lexical signs from the American Sign Language Lexicon Video Dataset (ASLLVD), including both 1- and 2-handed signs, we achieve a top-1 accuracy of 93.3% and a top-5 accuracy of 97.9%. The high probability with which we can produce 5 sign candidates that contain the correct result opens the door to potential applications, as it is reasonable to provide a sign lookup functionality that offers the user 5 possible signs, in decreasing order of likelihood, with the user then asked to select the desired sign.

Keywords: Sign Recognition, Model-based Machine Learning, Computer Vision, American Sign Language (ASL)

1. Introduction

Whereas many older approaches to computer-based sign recognition from video had focused on a selection of features known to be linguistically relevant to sign production, more recent research that has exploited neural nets has generally not attended to what is known about linguistic structure. The latter approaches do not work well, however, in the absence of large quantities of annotated data, quantities that exceed what is generally available for sign languages currently. Furthermore, they fail to provide insights into cases where the recognition fails.

To address the linguistic and computer vision complexities associated with automatic sign recognition, we have developed a novel hybrid approach that utilizes a set of known linguistic properties of the language to optimize the parameterization for state-of-the-art machine learning methods. These methods also rely on linguistically annotated data for citation-form signs from our American Sign Language Lexicon Video Dataset (ASLLVD) (Neidle, Thangali, and Sclaroff, 2012).[1]

Our 3-step approach differs from most other methods since it uses parameters related to upper body and hand and face configuration, coupled with linguistic constraints (as reflected in the statistics from the dataset).

1) We first use neural networks to automatically extract the 2D upper body and facial features from a signer's video sequence. These features are then used to estimate the 2D pose of the signer, and then, using dynamic programming, to fit a 3D model to estimate the related parameters. We also extract hand features using another neural net trained for handshape recognition.

2) We then introduce linguistic dependencies to adjust the probabilities of estimated start and end handshapes; these are based on precomputed co-occurrence probability priors for start/end handshape combinations. We also add a parameter related to the possible relationships between handshapes on the 2 hands in 2-handed signs.

3) The previously estimated parameters related to the upper body and handshape probabilities, modified with linguistically based information, are then used in a modified Hidden Conditional Ordinal Random Field (HCORF) for sign recognition.

This unified hybrid framework for sign recognition offers impressive sign recognition results in a fully scalable manner. Using a 350-sign vocabulary of isolated, citation-form lexical signs, we achieve a top-1 accuracy of 93.3% and a top-5 accuracy of 97.9%.

Section 2 briefly situates our current approach in the context of previous attempts at sign recognition. Section 3 presents our framework; the experiments and results are summarized in Section 4. In Section 5, we discuss possible applications of this technology.

2. Previous Achievements in Sign Recognition

In the early 2000's, isolated sign recognition from video or RGBD sensors, often using features of the signing known to be linguistically significant (e.g., Bowden et al., 2004), demonstrated some success on small vocabularies.

Signer independence poses additional challenges. Von Agris et al. (2006), using extracted image features, achieved 96.9% signer-independent recognition of 153 signs from 4 native signers of British Sign Language. Later, von Agris, Knorr, and Kraiss (2008), by combining 2D motion trajectories, facial features, and a hand model, achieved 88.3%, 84.5%, and 80.2% respectively for signer-independent recognition of vocabularies of 150, 300, and 450 signs from 25 native signers of German Sign Language. These results indicate that scalability is an issue.

Zaki and Shaheen (2011), using hand-crafted features describing handshape and orientation, place of articula-

[1] See http://www.bu.edu/av/asllrp/dai-asllvd.html. This dataset is also available at http://secrets.rutgers.edu/dai/queryPages/search/search.php and forms the basis for our new Web-accessible ASLLRP Sign Bank, accessible at http://dai.cs.rutgers.edu/dai/s/signbank (Neidle et al., 2018). The Sign Bank examples that were recorded as isolated signs, in citation form, are taken from the ASLLVD; the Sign Bank also includes additional examples taken from continuous signing.

tion, and hand motion, report 89.9% success in recognizing 30 ASL signs from 3 signers from the RWTH-BOSTON-50 database (Zahedi et al., 2005; that database is, in fact, comprised of a subset of 50 signs taken from the ASL data we had made publicly available and which are now shared through our Data Access Interface (DAI, and the new DAI 2); see Footnote 1).

For larger vocabularies, Cooper et al. (2011) attained 71.4% top-1 accuracy on a set of 984 signs from British Sign Language, but all from a single signer. Wang et al. (2016) achieved 70.9% accuracy on 1,000 isolated signs in Chinese Sign Language across multiple signers. However, they relied on an RGBD sensor for 3D information.

More recent approaches to sign language recognition, although focused on continuous signing rather than isolated signs, have been spurred by advances in neural nets. Such purely data-driven end-to-end approaches have been based on Recurrent Neural Net (RNN) architectures (e.g., Cui, Liu, and Zhang, 2017). Koller, Zargarin, and Ney (2017) use such an architecture, incorporating HMMs and 2D motion trajectories (but without integration of linguistic knowledge) to achieve 45.1% accuracy. Their multi-signer performance (27.1%) demonstrates that such methods do not generalize easily.

It is difficult to make direct comparisons with other sign recognition results because of vast differences in the nature of the data and conditions for research reported in the literature. In general, however, as the size of the dataset increases, the accuracy of isolated sign recognition has decreased. Methods used have not proved to be scalable. Our methods achieve both high accuracy in sign recognition on sizable vocabularies and scalability.

3. Overview of our Sign Recognition Framework

Our hybrid approach uses 1) discriminative neural net based computer vision methods coupled with generative methods for hand and pose feature extraction and related parameters, 2) additional linguistically driven parameters (Sections 3.1, 3.2), with enhancement of parameters from known linguistic dependencies (Section 3.3); and 3) scalable machine learning methods for sign recognition using the extracted parameters (Section 3.4); see Figure 2.

This results in improved sign recognition compared to previous approaches, because of the reduced parameterization and the efficiency of the algorithms, which are capable of coping with limited quantities of annotated data.

3.1 Summary of Features

Using the framework just described, we estimate a comprehensive set of features, with regard to: a) handshapes, b) number of hands, c) 3D upper body locations, movements of the hands and arms, and distance between the hands, d) facial features, and e) contact.

a) Features related to **handshape** are extracted from a neural net.

b) Signs are categorized based on the **number of hands** (1 vs. 2 hands) and the **degree of similarity of the handshapes** on the 2 hands for 2-handed signs.

c) The **upper body** parameters include 3D joint locations for the shoulders, arms, and wrists; velocities; and the **distance between the hands.**

d) The features for the **face** include 66 points (visible in Figure 1) from 3D estimates for the forehead, ear, eye, nose, and mouth regions, and their velocities across frames.

e) The **contact** parameters are extracted from our 3D face and upper body movement estimation, and relate to the possibilities of the hand touching specific parts of the body, e.g., the forehead or other parts of the face, arms, upper body, or the other hand.

The initial parameter values will, in some cases, be subsequently modified based on linguistic considerations, to be discussed in Section 3.3. This comprehensive set of parameters is then used within our CRF-based machine learning framework for purposes of sign recognition.

3.2 Feature Parameter Extraction

Next we describe how these parameters are extracted.

3.2.1 Upper Body, Hands, and Arms

We model upper body pose and use the 3D joint locations as features. We use Convolutional Neural Nets (CNNs) for initial estimation of 2D pose. We then apply a nearest neighbor matching coupled with a dynamic programming approach to search for the optimal 3D pose and part confidence maps (Dilsizian et al., 2016).

Using this 3D approach, we also extract linguistically important parameters, such as 3D motion trajectories, information about the number of hands (1- vs. 2-handed) and events involving contact between the 2 hands or contact with the face or body, as shown in Figure 1.

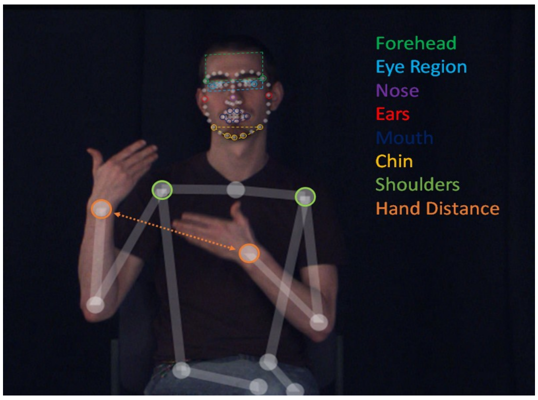


Figure 1. Locations where contact occurs

Handshape feature extraction and recognition have previously been demonstrated (Dilsizian et al., 2014; Ricco and Tomasi, 2009) with reasonable accuracy on limited datasets. More recently, CNNs have been used for robust recognition of New Zealand Sign Language handshapes from a large dataset with high variability (Koller, Ney, and Bowden, 2016). In our approach, we generate an ASL dataset for handshape recognition based on the publicly available ASLLVD corpus, which we plan to make available on the Web. We use wrist locations and forearm orientation to identify bounding boxes around the hands. We consider handshapes for which sufficient examples are available in the dataset (as is the case for 74 of the 86 handshapes). We balance the dataset by taking perturbations of shapes with fewer examples. Then we separate out 80% of all the obtained handshapes to be used as training exemplars in a CNN.

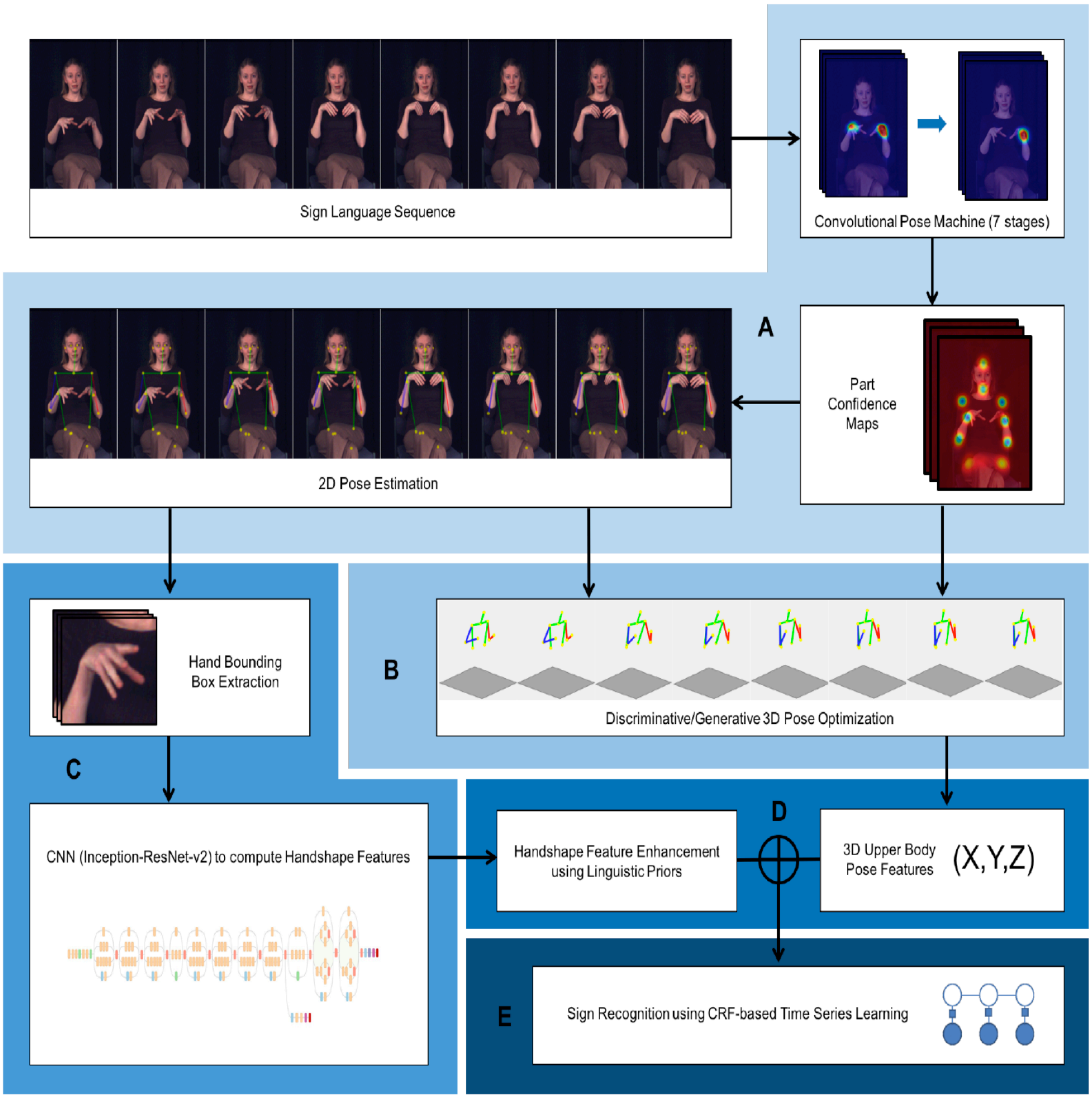


Figure 2. Overview of our sign recognition framework. (A) CNN-based 2D pose estimation, (B) novel discriminative/ generative 3D pose estimation, (C) CNN-based handshape recognition, (D) linguistic enhancement and feature combination, and (E) CRF-based sign recognition

The human annotations make use of a set of discrete handshape labels. However, hand configurations exhibit variations along a continuum (i.e., they are not discrete). In addition, actual handshape configurations produced in the course of signing frequently differ from the canonical handshapes we are using in our idealization, and even humans may have difficulty in determining which is the closest canonical handshape for a given realization. To capture the varying production of the handshapes that are key to sign identification (the start and end handshapes being the most informative), we consider the entire set of output probabilities (for each handshape) of the CNN to be features for sign recognition, rather than focusing on a single handshape label with the highest probability.

In order to capture a set of output probabilities that is sufficiently descriptive, we must avoid overfitting to prevent the CNN from converging entirely to the most probable handshape labels during the course of a sign. We train Inception-ResNet-v2 (Szegedy et al., 2017) on the hand images because of its ability to capture information from both local and global appearance.

Although we use the entire set of handshape output probabilities computed by our CNN as features for sign recognition, we report handshape prediction accuracy to demonstrate the effectiveness of our approach. We achieve a top-1 accuracy of 70.1% on the testing dataset after 20 epochs of training. The top-5 accuracy reaches 92.3%. The top-1 and top-5 accuracy for the test set is shown over training epochs in Figure 3.

Thus, in the initial phase of our handshape feature extraction, we compute a vector of handshape probabilities (with a length of 74, as we are using the 74 handshapes for which we have a sufficient number of examples) for each hand in each frame during the production of signs in our sign recognition dataset.

3.2.3 Face and Head

Non-manual features have been shown to improve recognition of manual signs (von Agris, Knorr, and Kraiss, 2008; Koller, Forster, and Ney, 2015). Thus we estimate the 3D locations of 66 points on the face, as well as head movement, to include all possible informative non-manual information.

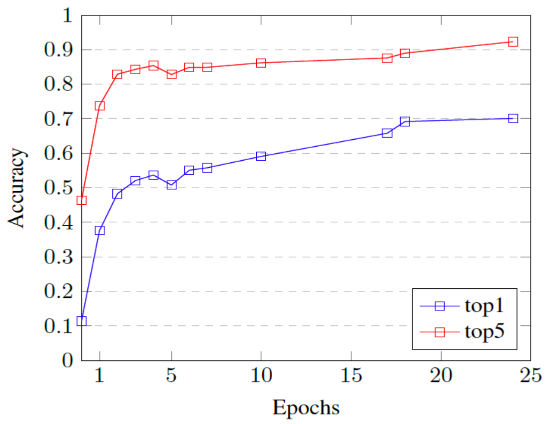


Figure 3. Top-1 and top-5 handshape recognition accuracy on test set by training epoch.

3.3 Incorporation of Linguistic Modeling for Enhancement of Parameter Estimates

The initial estimates of several of the above parameters can be refined based on known linguistic dependencies.

3.3.1 Dependencies between Start & End Handshapes

We exploit phonological constraints that hold between start and end handshapes in lexical signs to refine the handshape estimates for start and end handshapes (Thangali et al., 2011; Thangali 2013; Dilsizian et al., 2014). These dependences are reflected in the co-occurrence probabilities from our dataset.

3.3.2 Dependencies between Dominant & Non-dominant Handshapes in 2-handed Signs

We distinguish 2-handed signs that have essentially the same handshape on both hands from those that involve different handshapes, based in part on the handshape similarity parameter mentioned earlier. In the former case, we can boost handshape accuracy by combining information from the independent handshape estimates for the 2 hands. In the latter case, handshape possibilities for the non-dominant hand are significantly constrained.

3.4 Sign Recognition

We use the above extracted parameters as input to a structured Conditional Random Field (CRF) method—a modified Hidden Conditional Ordinal Random Field (HCORF) (Walecki et al., 2015)—to recognize signs. In addition, for each sequence, our modified HCORF includes an additional error term that measures the error between start/end handshape predictions and ground truth labels.

The advantages of our linguistically motivated, reduced parameter approach are demonstrated in the next section.

4. Sign Recognition Experiments and Results

4.1 Dataset

In this research we focus on lexical signs, the largest morphological class of signs. For training, we used the most comprehensive publicly accessible, linguistically annotated, video collection of isolated ASL signs, the American Sign Language Lexicon Video Dataset (ASLLVD) (Neidle, Thangali, and Sclaroff, 2012); see also Footnote 1. The ASLLVD itself includes over 8500 examples corresponding to almost 2800 monomorphemic lexical signs in citation form from 6 native signers. However, for these experiments, we selected a set of 350 signs from among those that had the greatest number of examples and signers per sign. On average, there were 4.7 signers and 6.9 total examples per sign for this set of 350 signs (a total of about 2400 examples). This was sufficient to train our neural nets.

4.2 Experiments

For each frame in each video sequence, we extract a feature vector of dimension 110, which includes the previously discussed features (handshape, motion trajectory, and other linguistically motivated features). This feature vector is used as input to our machine learning framework for sign recognition. We trained on our dataset, which generally contained 4-6 signers per example, using 80% of the data for training and 20% for testing. For each sign, 2 examples were randomly selected to be in the testing set, and the remaining examples were used for training. We tested on vocabularies of differing sizes (175 vs. 350 signs) to test the efficiency and scalability of our approach. We also performed a series of experiments to separate out the contributions of the different parameters.

4.3 Results

As shown in Figure 4, from a vocabulary of 350 signs (including both 1- and 2-handed signs), using all of our parameters, we achieve a top-1 accuracy of 93.3% and a top-5 accuracy of 97.9%. Figure 4 demonstrates the advantage of: 3D pose over 2D (green vs. amber); the addition of contact parameters (red); and the inclusion of all linguistic parameters and constraints in our framework (blue).

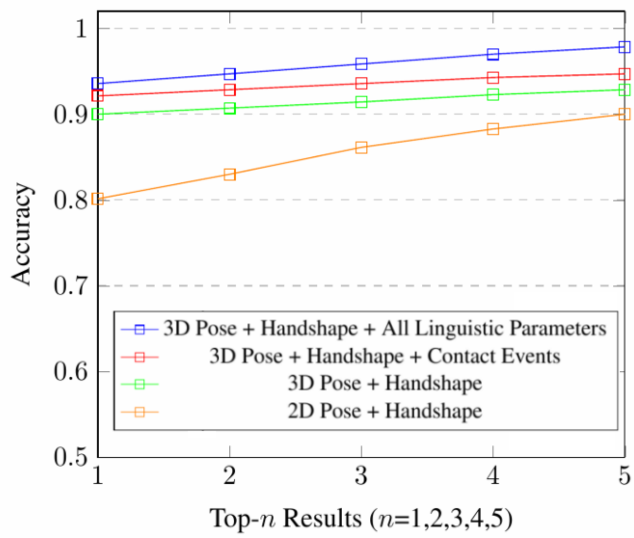


Figure 4. Contribution of Parameters to Accuracy

Comparing the results of vocabularies of 175 vs. 350 signs (Figure 5), accuracy declines by only 2.1% for top 1, and by only 1.3% for top 5 with the larger vocabulary. This provides evidence for the scalability of the approach.

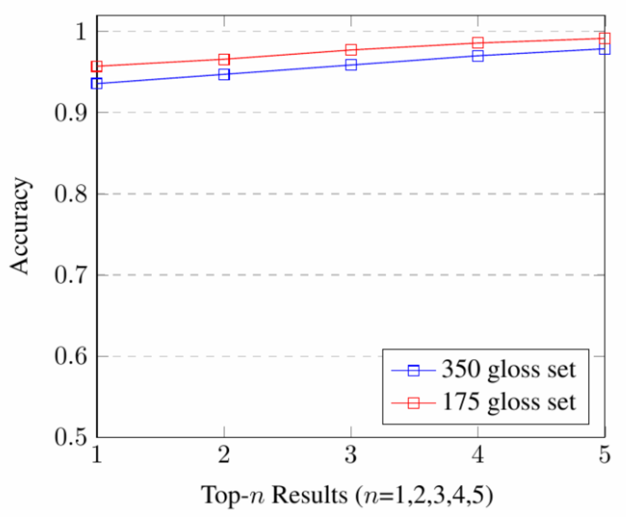


Figure 5. Comparing the Results on Vocabularies of 175 vs. 350 Signs

5. Significance for Potential Future Applications

There are many possible practical applications of technology for sign identification from video. For example, sign lookup capability would present significant benefits to Deaf communities, and to others wanting access to sign language resources such as dictionaries. Sign language dictionaries are currently often accessed by means of the written language, e.g., looking up a sign in an ASL resource by searching for a possible English translation of that sign. This has obvious drawbacks, as the user (whether Deaf or hearing) may not know the corresponding word from the spoken/written language. Available alternatives, which are in use for some sign language resources, generally involve laboriously having the user specify multiple features of the sign, such as handshape, location, movement type; this constitutes a very inefficient and unsatisfying lookup mechanism.

Our goal is to develop a lookup functionality that would enable users to search through our own electronic resources (Neidle et al., 2018), or to use our lookup interface to access other resources, through one of two input methods: either by producing the target in front of a webcam, or by identifying the start and end frames of the sign of interest from a video with continuous signing.

Although additional research will be required before such a lookup mechanism can be provided, the fact that we currently achieve about 98% success, using scalable methods, in identifying five candidate signs that include the target sign is extremely encouraging. It would be practically reasonable to offer the user 5 choices, in decreasing order of likelihood, as part of the lookup process, with the user able to view those sign videos and choose among the signs before confirming the selection and proceeding with the lookup, as sketched in Figure 6. Final design of such an interface will also involve consultation with prospective users of such tools.

6. Conclusions

We have demonstrated a general framework for recognition of isolated signs produced by multiple signers. Our framework leverages linguistic structure and dependencies, thereby enabling it to work from limited quantities of annotated data and to outperform previous methods. Our parameter extraction methods are based on state-of-the-art 3D handshape, face, and upper body parameter estimation, as well as integration of linguistic properties and constraints. The resulting modified parameter vector allows for a scalable and efficient approach to sign recognition.

In the future, we plan to expand the corpus and associated annotation sets to further improve the performance of our methods. We also intend to refine/augment the linguistically motivated features to enhance recognition accuracy, which would not be possible with purely data-driven methods. Furthermore, the methods being developed will, we hope, have beneficial practical applications, which we intend to pursue.

Figure 6. Example of Choices to be Offered to a User before Confirmation of a Lookup Selection – Based on the Interface for our Current ASLLRP Sign Bank

7. Acknowledgments

This research was partially funded by grants from the National Science Foundation (#1748016, 1748022, 1059218, 0705749, 0958247, 1703883, 1567289, 1555408, 1451292, 1447037). We also gratefully acknowledge contributions by many students and colleagues at Boston, Gallaudet, and Rutgers Universities and we thank our many ASL linguistic consultants. Special thanks to Stan Sclaroff, Ashwin Thangali, Joan Nash, and Vassilis Athitsos for their participation in data collection efforts for the ASLLVD, the dataset that served as the basis for the research reported here, and to Christian Vogler and Augustine Opoku for their help in developing our Web interfaces for data sharing.

8. Bibliographical References

Bowden, R., Windridge, D., Kadir, T., Zisserman, A. and Brady, M. (2004). A Linguistic Feature Vector for the Visual Interpretation of Sign Language. Proceedings of the ECCV.

Cooper, H., Holt, B. and Bowden, R. (2011) Sign Language Recognition. In Moeslund, T. B., Hilton, A., Krüger, V. and Sigal, L., (eds.) *Visual Analysis of Humans: Looking at People*: Springer. pp. 539–562.

Cui, R., Liu, H. and Zhang, C. (2017). Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization. Proceedings of the CVPR 2017, Honolulu, Hawaii.

Dilsizian, M., Tang, Z., Metaxas, D., Huenerfauth, M. and Neidle, C. (2016). The Importance of 3D Motion Trajectories for Computer-based Sign Recognition. Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining. LREC 2016, Portorož, Slovenia. May 2016.

Dilsizian, M., Yanovich, P., Wang, S., Neidle, C. and Metaxas, D. (2014). A New Framework for Sign Recognition based on 3D Handshape Identification and Linguistic Modeling. Proceedings of the LREC 2014, Reykjavik, Iceland. May 2014.

Koller, D., Ney, H. and Bowden, R. (2016). Deep Hand: How to Train a CNN on 1 Million Hand Images when your Data is Continuous and Weakly Labelled. Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Koller, O., Forster, J. and Ney, H. (2015) Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding, 141*, pp. 108-125.

Koller, O., Zargaran, S. and Ney, H. (2017). Re-Sign: Re-Aligned End-To-End Sequence Modelling With Deep Recurrent CNN-HMMs. Proceedings of the CVPR 2017, Honolulu, Hawaii.

Neidle, C., Opoku, A., Dimitriadis, G. and Metaxas, D. (2018). NEW Shared & Interconnected ASL Resources: SignStream® 3 Software; DAI 2 for Web Access to Linguistically Annotated Video Corpora; and a Sign Bank. Proceedings of the 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community. LREC 2018, Miyagawa, Japan. May 2018.

Neidle, C., Thangali, A. and Sclaroff, S. (2012). Challenges in Development of the American Sign Language Lexicon Video Dataset (ASLLVD) Corpus. Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. LREC 2012, Istanbul, Turkey. May 2012.

Ricco, S. and Tomasi, C. (2009) Fingerspelling Recognition through Classification of Letter-to-Letter Transitions. *9th Asian Conference on Computer Vision, Xi'an, China*.

Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A. A. (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. *AAAI*, pp. 4278–4284.

Thangali, A. (2013) *Exploiting Phonological Constraints for Handshape Recognition in Sign Language Video*. Unpublished, Doctoral Dissertation, Boston University.

Thangali, A., Nash, J. P., Sclaroff, S. and Neidle, C. (2011). Exploiting Phonological Constraints for Handshape Inference in ASL Video. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) 2011.

von Agris, U., Knorr, M. and Kraiss, K. F. (2008). The significance of facial features for automatic sign language recognition. Proceedings of the International Conference on Automatic Face & Gesture Recognition.

von Agris, U., Schneider, D., Zieren, J. and Kraiss, K.-F. (2006). Rapid signer adaptation for isolated sign language recognition. Proceedings of the Workshop on Vision for Human Computer Interaction (V4HCI).

Walecki, R., Rudovic, O., Pavlovic, V. and Pantic, M. (2015). Variable-state latent conditional random fields for facial expression recognition and action unit detection. Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2015, IEEE.

Wang, H., Chai, X., Hong, X., Zhao, G. and Chen, X. (2016) Isolated Sign Language Recognition with Grassmann Covariance Matrices. *ACM Transactions on Accessible Computing (TACCESS), 8*.

Zahedi, M., Keysers, D., Deselaers, T. and Ney, H. (2005) Combination of Tangent Distance and an Image Distortion Model for Appearance-Based Sign Language Recognition In *Pattern Recognition*, Berlin / Heidelberg: Springer. pp. 401-408.

Zaki, M. M. and Shaheen, S. I. (2011) Sign Language Recognition using a Combination of New Vision Based Features. *Pattern Recognition Letters, 32*(4), pp. 572-77.

Workflow Management and Quality Control in the Development of the PJM Corpus: The Use of an Issue-Tracking System

Piotr Mostowski, Anna Kuder, Joanna Filipczak, Paweł Rutkowski
Section for Sign Linguistics, Faculty of the Polish Studies, University of Warsaw,
Krakowskie Przedmieście 26/28, 00-927 Warsaw, Poland
{piotr.mostowski, anna.kuder, j.filipczak, p.rutkowski}@uw.edu.pl

Abstract

The main goal of the present paper is to describe a workflow management and quality assurance system used in the project of developing the Polish Sign Language (*polski język migowy*, PJM) Corpus currently underway at the University of Warsaw, Poland. To ensure a satisfactory level of annotation quality, we implemented an external issue-tracking system as a basic tool to manage all stages of the annotation process: segmenting the video recording into individual signs, adding glosses to the delineated signs, segmenting text into clauses, translating text into written Polish and adding grammar tags marking different language phenomena. This paper offers a detailed overview of the procedures that we employ, illustrating the most important advantages and disadvantages of our approach and the choices we have made.

Keywords: sign language, corpus linguistics, corpus building, annotation, tracking system, quality control management

1. Introduction

The Polish Sign Language (PJM) Corpus, which is currently being developed at the University of Warsaw's Section for Sign Linguistics (UW SSL)[1], ranks among the largest sign language corpora that are being created worldwide. It was inspired by the development of other such projects, including the Australian Sign Language (Auslan) corpus[2] (Johnston, 2009) the Dutch Sign Language (NGT) corpus[3] (Crasborn and Zwitserlood, 2008), the British Sign Language (BSL) corpus[4] (Schembri et al., 2013) and the German Sign Language (DGS) corpus[5] (Hanke et al., 2010). The main idea behind the PJM Corpus project is to collect a large set of video clips showing Polish Deaf signers using PJM in different contexts. Even though work on the corpus is not finished (the project was launched in 2010 and will continue until at least 2019), it is already being used for a range of different purposes, which include: conducting linguistic research, studying Deaf culture, enhancing the qualifications of PJM teachers and interpreters, compiling dictionaries and carrying out comparative studies between sign languages.

2. Building a Sign Language Corpus

The process of building a sign language corpus, a tremendously labor-intensive task, can be divided into two main phases: obtaining a video archive of deaf people signing and annotating it. The first phase is usually accomplished via a number of recording sessions that take the form of filming a meeting of two deaf informants, who sit facing each other and respond to elicitation materials shown to them on a screen in a multi-media presentation (see, e.g., Hanke et al., 2010; Rutkowski et al., 2017). The raw material obtained in recording sessions is backed up, compressed and uploaded into special software, where it is then subject to linguistic processing.

For this purpose the UW SSL team uses the iLex software, developed at the University of Hamburg (Hanke and Storz, 2008). Another popular program used for this purpose is ELAN (Crasborn and Sloetjes, 2008). iLex, however, allows video materials and annotation files to be stored in the form of a single database that can be accessed online by many people at the same time. All changes implemented in the software are immediately visible to all of its users. ELAN, on the other hand, requires its users to work on corpus material locally on their computers. As the UW SSL annotation team consists of more than 20 people and the implemented annotation process is non-linear in its nature, it is more convenient to work in one database that can be accessed by many people simultaneously, hence the decision to use iLex for the PJM Corpus.

As of 2017, 134 Deaf informants have been recorded for the purposes of the PJM Corpus. Each recording session lasts approximately 4-5 hours. So, for the time being, this has resulted in approximately 600 hours of raw HD video material.

The second phase of building a corpus involves transforming the archive into a searchable database (e.g., Johnston, 2010). In order to accomplish this aim, researchers need to add different layers of linguistic information to the raw video data through the process of annotation. Annotating a sign language corpus is an extremely time-consuming task and can be done by humans only. There are no automatic or semi-automatic tools available and standards and good practices are only now being developed. As annotating requires language proficiency at the maximum level, the PJM Corpus is annotated only by Deaf or CODA signers. Hearing annotators with linguistic education only help with the methodological distinctions and in doubtful cases (Rutkowski et al., 2017).

The PJM Corpus is annotated on several different levels. After a recording session is first uploaded into iLex, it is given a specific name (e.g., 'K04AF01-11', 'K04AF12-16') and metadata is added to it in line with the annotation schema. Then this recording, now called a transcript, is segmented into more than 20 short video clips corresponding to the individual tasks performed by the informants during the recording session. After this is finished, the recording is subject to the annotation process, which, again, consists of a few steps.

[1] www.plm.uw.edu.pl/en
[2] www.elar.soas.ac.uk/Collection/MPI55247
[3] www.ru.nl/corpusngten/about-corpus-ngt/latest-news/
[4] www.bslcorpusproject.org/project-information/
[5] www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/the-project.html

First, annotators watch each clip separately and segment the stream of signs into individual tokens. This is an extremely time-absorbing process, with even a skillful annotator needing approximately one hour of work in order to segment one minute of continual signing into individual signs. Then each sign token is lemmatized and marked as an instance of a particular gloss. Annotators also mark signs that do not possess a clear linguistic status and are rather purely conversational, such as different kinds of gestures and palm-ups. After that the text is segmented into clauses and translated into written Polish. The annotation process finishes with the text being tagged with respect to a number of grammar parameters, which include:

- parts of speech;
- non-manual elements (head movements);
- non-manual elements (body movements);
- mouthing;
- repetition;
- word order;
- negation;
- argument structure and macro role structure.

During the process, quality control is performed twice: once after adding glosses to the individual sign occurrences and once after translation. Glossing work is overseen by a "superannotator" – a Deaf person with broad experience who has worked on the project from its beginning and is highly competent in the annotation guidelines for glossing. The superannotator is selected by a decision of the whole annotation team. This person's role is extremely important for ensuring annotation quality but also for positively impacting the work of the whole team. Oversight of the written translations, in turn, is performed by a skillful interpreter who works with the Deaf on a daily basis and is fluent in both PJM and Polish.

The annotation workflow described above is the outcome of a few years of continual work on the corpus and creating guidelines for annotation. It highlights how time-consuming the annotation process of the PJM Corpus is. Each video clip is inspected several times by different team members, each of them looking for and marking varied language phenomena. Different people segment, gloss, translate and tag the data. This process is non-linear in the sense that separate annotation stages are performed simultaneously on different parts of the material. We are positive that this is the only way of providing a fully annotated corpus that will be useful for research purposes. However, with a team as large as over 20 people working in locations all over the country, it would be impossible to complete this task without some centralized management tool to help avoid confusion and ensure actual growth of the annotated dataset. This was the main reason we decided to look for a convenient online managing system that could be helpful in this regard.

3. An Issue-Tracking System for Annotation Quality Management

3.1 YouTrack

In order to maintain control over the described annotation process, the UW SSL team implemented an existing issue-tracking system as a basic tool to manage all the work done in the project. We decided on YouTrack[6], an external tool developed by the software company JetBrains[7], which was chosen in part because it offered a free subscription for open source projects, which we acquired back in 2012. The rest of the present paper will be devoted to describing the solutions applied in YouTrack, although we are positive that the same can be achieved using any other popular issue tracker, for example Plutora[8], BugZilla[9], Backlog[10], JIRA[11] or RedMine[12].

YouTrack is an online bug and issue-tracking system used mainly by programmers or other specialists working in IT. Its main feature is the ability to create individual "issues" (each issue corresponds to one task that needs to be completed – in our case a given task from a given transcript in the corpus) with fully customizable fields, which determine all of the issue characteristics. The issues can be grouped, forming different, independent projects. YouTrack offers a user-friendly tool for searching for specific issues without having to know or use any programming language. It is possible for the project manager to easily create reports, use agile boards (designed to help teams plan and visualize their work through a special system of cards updated in real time), manage work time and control the work on many different levels within this system. Furthermore, there is an application for both iOS and Android which makes it possible to manage YouTrack projects from a mobile device.

3.2 Workflow in YouTrack

Using an issue-tracking system is straightforward and very helpful in large-scale projects like corpus annotation, but only after ensuring that the user knows exactly what she wants to accomplish. This means that the first important step is planning and creating the design of the whole workflow. As all the issue fields in the tracker are fully customizable, the possibilities it gives in designing the workflow are almost endless. However, the tracker would not be of much use if its user did not decide what steps should be undertaken and completed in order to accomplish the desired aim (in our case: full annotation of signed texts on all of the mentioned levels). The more fixed and fewer changeable points in the workflow, the greater the likelihood of the work running smoothly. The greatest advantage of using a tracking system lies in automating part of the work on the project, but in order to make use of this the work needs to be planned in great detail before it even starts.

The process of workflow design therefore precedes creating any project in YouTrack. This process consists in deciding on the issue template (what fields will be used and for what purposes), determining what stages will need to be performed in order for a task to become resolved and assigning appropriate users to the project. Only after the workflow is programmed can the project manager start creating issues within it.

The UW SSL uses YouTrack to manage work in a number of its research projects. It was used for controlling the work of the team creating the first *Corpus-based Dictionary of Polish Sign Language*[13] (Łacheta et al., 2016) and is also

[6] www.jetbrains.com/youtrack/features/issue_tracking.html
[7] www.jetbrains.com
[8] www.plutora.com
[9] www.bugzilla.org
[10] www.backlog.com
[11] www.atlassian.com/software/jira
[12] www.redmine.org
[13] www.slownikpjm.uw.edu.pl/en

employed in a few smaller projects. YouTrack is the most helpful, however, in managing the PJM Corpus annotation, as this project requires the most elaborate workflow involving the most numerous team. We find this issue tracker extremely useful in multistage, hierarchical projects.
For the PJM Corpus annotation project, the workflow in YouTrack was designed to mirror the workflow implemented in the annotation process described in the section 2 of the present paper. It is depicted in graphical form in Figure 4.
Each issue in the project corresponds to a single video clip from a given transcript in the iLex software. YouTrack gives an ID (e.g., 'NPRH-837') to each issue automatically, based on the name of the whole project (in our case: 'NPRH' – an acronym for the name of the research grant that financially supports corpus annotations). An individual issue's name is inserted manually in the appropriate field and, in our case, consists of a number of the task and a number of the corresponding transcript from the iLex software (see Figure 1 – *zadanie* means 'task' in Polish).

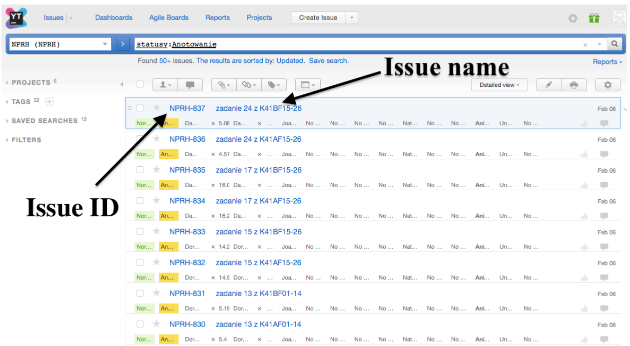


Figure 1: Issue list on the UW SSL's YouTrack webpage.

Each issue has 17 individual fields where all of the information about it are inserted (see Figure 2). In those fields we specify:

- priority of the task;
- actual status;
- current assignee;
- duration of the task (in minutes and seconds);
- annotator's name;
- the deadline for providing annotation;
- superannotator's name;
- clause tagger's name;
- the deadline for providing clause segmentation;
- interpreter's name;
- the deadline for providing written translation;
- the names of translation quality supervisors;
- PoS tagger's name;
- negation tagger's name;
- additional taggers' names.

Issues are created by the project manager, their fields filled out and issues are assigned to the appropriate team members. Each annotator has her own account in YouTrack with her specific roles and access. After logging in and clicking the 'assigned to me' button each person can see a list of all of her current tasks. Then she marks the tasks that she is currently working on by changing the issue status in the corresponding field. Then she logs into iLex and works on her clips. After her work is finished she changes the status of the issue in question in YouTrack and the value in the 'assignee' field automatically changes to the next responsible annotator's name. The next person then gets an automatic e-mail notification about a new task in her account and, after logging in, sees the issue on her 'assigned to me' list. Consecutive statuses marked with different colors (see Figure 3) correspond to the annotation stages of each corpus task mentioned in section 2 of the present paper. The 'assignee' field is programmed to change when the value in the 'status' field changes. All of the changes in the issue's fields are saved in its special bookmark called 'history' and are accessible anytime, which eliminates anonymity in the project and provides an easy way to control who is responsible for which alterations.

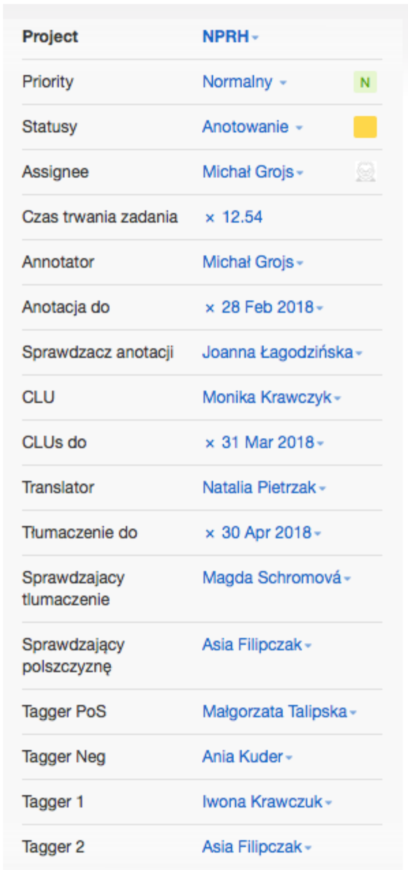


Figure 2: Example of an individual issue field list.

In the issue workflow we distinguish 11 kinds of status:

- annotation;
- checking of the annotation;
- clause segmenting;
- translating;
- checking of the translation (2 times);
- PoS tagging;
- negation tagging;
- additional tagging (2 times),
- issue is resolved.

Each stage-completed status (marked with a color) is paired with a corresponding status stating that the work on that stage is currently underway (without any color, marked as *w trakcie*, Polish for 'underway', in Figure 3).

Figure 3: List of all the issue status types.

If at any stage the superannotator should come across any mistakes in the annotation that ought to be corrected by the original annotators, they are allowed to 'break' the workflow and direct the task in question back to the team member responsible.
Each task has its individual space for comments, which is used for discussions between the users whenever any problems or disagreements appear. Users can tag each other in the comments and by so doing send each other e-mails with the comments.
Furthermore, YouTrack enables periodic report-generation. The straightforward way of generating reports stating how many tasks were completed and by whom in a given period of time is an invaluable help in creating different kinds of summaries, e.g., for grant-acquiring purposes. The UW SSL YouTrack project manager also counts the time of the annotated issues (from the appropriate fields) to keep track of how much video corpus material is already fully or partially annotated. This allows the achieved work progress to be monitored and compared against the scheduled milestones, the work of annotators to be periodically assessed and provides a basis for calculating the annotators' salaries.
All of the team members, Deaf and hearing, use YouTrack on a daily basis. When a new person joins the team she gets accounts in both iLex and YouTrack and is trained in using both tools simultaneously.

4. Advantages and Disadvantages of Using an Issue Tracker in a Large-Scale Project

The UW SSL team has been using YouTrack continually since late 2012. After five years of working with this tool we have become aware of many of its advantages and drawbacks affecting the team administration and work management, and we will share these observations here.

Firstly, designing a project in YouTrack forces every aspect of the work to be planned before it even starts. This helps in prioritizing some work stages over others and building the logical, hierarchical structure of the workflow in order to accomplish the desired aim.
Moreover, using an issue tracker facilitates the management of an extremely broad and rapidly growing set of tasks. The project manager can search in seconds for issues that are interesting from some particular point of view, check the status of a given issue at a given time or generate a report on the work done in the project. This makes the whole work done in the project more transparent. It is important that everyone knows what each user is supposed to work on, but also was she has done in the past. The lack of anonymity can positively impact the quality of annotations.
The tracker allows the work of all the users to be monitored, as it shows the date of the last login and of recently applied changes. It is easy to react when an annotator is working more or less than she ought to.
YouTrack helps ensure that the annotation process is done consistently and that each task undergoes the same stages before it is resolved. It also allows non-linear work – any task can be accessed at any time and annotators are not obliged to wait for their colleagues to complete their work in order to start theirs. Everyone can work simultaneously on different parts of material.
If the workflow is designed and programmed properly, YouTrack guarantees automation of part of the work that does not require human involvement, thus saving valuable time and costs.
Overall, the tracker interface is transparent and user-friendly. Despite some initial reluctance, all members of the team learned how to use the tool very quickly and by now there are virtually no problems with operating the system.
However, there are still some potential drawbacks when it comes to using YouTrack. One is that the team has to have a person responsible for operating the platform, who will act as the project manager. This person has to create projects within YouTrack, program their workflow, add new users, assign their roles and generate reports.
The system also requires someone responsible for creating all of the issues, filling out their fields and assigning them to appropriate users. This can be done by hand only, but in some cases one command can be applied to a whole set of issues at the same time (speeding up the work). This is without question the most time-consuming task in using an issue tracker but is relatively straightforward and easy to learn.
As most people are not familiar with using an issue tracker on a daily basis, it can be overwhelming for new members of the team at first sight. This is why training is required when a new person joins the team. It is also advisable to assign the least possible access to new YouTrack users before they get comfortable with using the tool.
Once the workflow is designed, the issues are already created and the work has begun, it is not easy to implement any changes in the project. To overcome this, it is advisable to start by creating a test project, which can be evaluated by the users and only after receiving their feedback to design and create the final issue-tracking project.

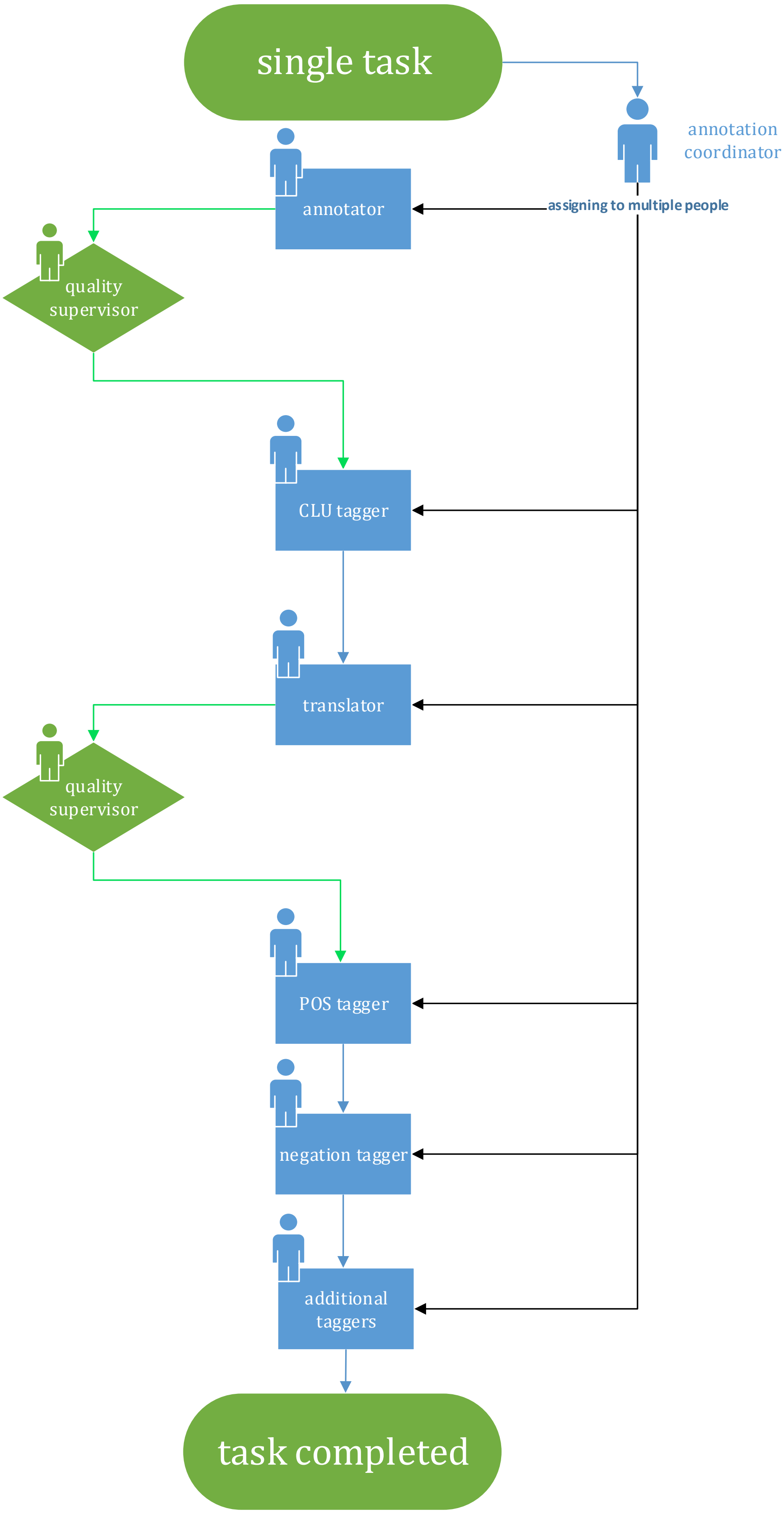


Figure 4: Workflow of each tasks in SSL's YouTrack.

5. Conclusions

The UW SSL team applied all of the solutions described in the present paper in order to simplify the daily work of a large group of people, who work in different cities and at different times. The fact that both YouTrack and iLex are accessible online makes it possible to work on the PJM Corpus annotation anytime and anywhere (with web access). It would be virtually impossible to control the work of such a large group in any decentralized way (e.g. only through e-mails or using some spreadsheet program).

The system seems to be working very well, as in the annotation process the annotators, translators and taggers have so far identified 5,500 different lexemes (which have been divided into 14,200 sublexemes), glossed more than 504,000 individual sign tokens, translated more than 10,000 PJM clauses into Polish sentences and tagged approximately 100,000 tokens for their grammatical features.

The SSL team uses YouTrack extensively, not only for managing the annotation process itself, but also, as mentioned above, in the creation of the *Corpus-based Dictionary of Polish Sign Language* (Łacheta et al., 2016) and several smaller projects. Each of those research projects has its own corresponding project in the tracking system with customized fields – each of the employed workflows was designed from scratch so as to best suit the team's needs.

In this paper, we have listed what are, in our experience, the main advantages and potential drawbacks of using an issue-tracking system. Overall, however, we strongly encourage any large research team to use this or a similar tool to simplify their workflow, which will lead to more efficient and carefree work.

6. Acknowledgements

The first phase of the PJM Corpus project was supported financially by Poland's National Science Center (*Narodowe Centrum Nauki*) under the project *Iconicity in the grammar and lexicon of Polish Sign Language* (*PJM*) (grant number: 2011/01/M/HS2/03661) and by the Foundation for Polish Science (*Fundacja na rzecz Nauki Polskiej*) under the project *Grammatical categorization through space and movement in Polish Sign Language* (grant number: 1/2009). The second phase is currently financed by the Polish Ministry of Science and Higher Education (*Ministerstwo Nauki i Szkolnictwa Wyższego*) under the National Program for the Development of Humanities (*Narodowy Program Rozwoju Humanistyki* – project title: *Multi-layered linguistic annotation of the corpus of Polish Sign Language* (*PJM*); grant number: 0111/NPRH3/H12/82/2014).

7. Bibliographical References

Crasborn, O. & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. 6th International Conference on Language Resources and Evaluation* (*LREC'08*). Paris: ELRA, pp. 39–43.

Crasborn, O. & Zwitserlood, I. (2008). The Corpus NGT: an online corpus for professionals and laymen. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. 6th International Conference on Language Resources and Evaluation* (*LREC'08*). Paris: ELRA, pp. 44–49.

Hanke, T. & Storz, J. (2008). iLex – A database tool for integrating sign language corpus linguistics and sign language lexicography. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. 6th International Conference on Language Resources and Evaluation* (*LREC'08*). Paris: ELRA, pp. 64–67.

Hanke, T., König, L., Wagner, S. & Matthes, S. (2010). DGS Corpus & Dicta-Sign: The Hamburg Studio Setup. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. 7th International Conference on Language Resources and Evaluation* (*LREC'10*). Paris: ELRA, pp. 106–109.

Johnston, T. (2009). Creating a corpus of Auslan within an Australian national corpus. In *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages*. Somerville: Cascadilla Proceedings Project, pp. 87–95.

Johnston, T. (2010). From archive to corpus: transcription and annotation in the creation of signed language corpora. *International journal of Corpus Linguistics,* 15(1), pp. 97–114.

Łacheta, J., Czajkowska-Kisil, M., Linde-Usiekniewicz, J. & Rutkowski, P. (Eds.) (2016). *Korpusowy słownik polskiego języka migowego/Corpus-based Dictionary of Polish Sign Language*, Warsaw: Faculty of Polish Studies, University of Warsaw.

Rutkowski, P., Kuder, A., Filipczak, J., Mostowski, P., Łacheta, J. & Łozińska, S. (2017). The design and compilation of the Polish Sign Language (PJM) Corpus, In P. Rutkowski (Ed.), *Different faces of sign language research*. Warsaw: University of Warsaw, Faculty of Polish Studies, pp. 125–151.

Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S. & Cormier, K. (2013). Building the British Sign Language Corpus. *Language Documentation & Conservation,* 7, pp. 136–154.

YouTrack documentation: https://www.jetbrains.com/youtrack/features/issue_tracking.html

Per Channel Automatic Annotation of Sign Language Motion Capture Data

Lucie Naert, Clément Reverdy, Caroline Larboulette, Sylvie Gibet
IRISA Lab., Université Bretagne Sud
Vannes, France
{lucie.naert, clement.reverdy, caroline.larboulette, sylvie.gibet}@univ-ubs.fr
lsf.irisa.fr

Abstract

Manual annotation is an expensive and time consuming task partly due to the high number of linguistic channels that usually compose sign language data. In this paper, we propose to automatize the annotation of sign language motion capture data by processing each channel separately. Motion features (such as distances between joints or facial descriptors) that take advantage of the $3D$ nature of motion capture data and the specificity of the channel are computed in order to (i) segment and (ii) label the sign language data. Two methods of automatic annotation of French Sign Language utterances using similar processes are developed. The first one describes the automatic annotation of thirty-two hand configurations while the second method describes the annotation of facial expressions using a closed vocabulary of seven expressions. Results for the two methods are then presented and discussed.

Keywords: automatic annotation, automatic segmentation, motion capture, French Sign Language, linguistic channels

1. Introduction

Whether we want to linguistically study sign languages, use digital data to identify salient linguistic components, recognize or synthesize continuous signing, an annotation of the data is needed. The annotation of sign language is a two-step process. The first step, called *segmentation*, consists in dividing the stream of data in segments of interest. Those segments are then identified in a second step called *labeling*. This annotation might be done manually but is a fastidious, time consuming and expensive task. Not only does it require the skills of language experts, but it is also subject to inaccuracies and mistakes as the experts may not have exactly the same segmentation criteria. In the particular case of sign language, the annotation process needs to be done by experts in sign language and gesture annotation. In addition, sign languages are expressed simultaneously on multiple channels (manual configuration, wrist orientation, body posture, facial expression, etc.), thus complicating the task of the annotators. When comparing the duration of the annotation process to the duration of the data to annotate, (Dreuw and Ney, 2008) introduces a real-time factor of 100 (i.e. all the manual and non manual features of a 1 minute video of sign language will be annotated in 100 minutes). We propose to automatically annotate each channel separately following the scheme of Figure 1.

Automatic annotation of sign languages could reduce the annotation costs but is still a challenging and yet to be solved task. One way to tackle this problem is the *machine learning* approach (e.g. use of Bayesian/statistical models or artificial neural networks) which aims at automatically learning the parameters of a model from a sample of manually annotated data. This model is then used to automatically annotate new data. The corpus intended to train the machine learning model must thus be designed carefully before recording sign language data using either video or motion capture technologies. Despite being easy to use and relatively cheap, video does not preserve the third spatial dimension of motion. Furthermore, the resolution of classic video recordings is rarely high enough to obtain a precise segmentation. Motion capture (MoCap) technologies offer

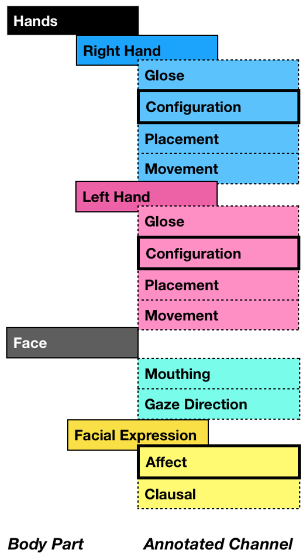


Figure 1: Our intended annotation scheme: currently, we automatically annotate the hand configuration and affect channels (boxes with bold borders).

a better precision, both spatially and temporally, and spatial information that is not available with 2D cameras. MoCap data can be used for sign language analysis – using motion descriptors computed from the 3D data, such as distances between joints, velocity or curvature of selected joints – as well as for synthesis by using the captured motions to animate a signing avatar. In this paper, we will address the problem of the automatic annotation of sign language MoCap data, focusing more specifically on two main channels: facial expression, and hand configuration.

Previous work on automatic annotation on video and MoCap data are described in section 2. The specification, capture and manual annotation of the input sign language data used in this study is introduced in section 3. In section 4, the annotation methodology is presented and illustrated with the examples of the automatic annotation of the hand configurations and the facial expressions in French Sign Language. Section 5 describes the results of the automatic annotation applied to those two channels. Finally, section 6 discusses the perspectives and challenges of our method.

2. Related Work

In this section, previous studies on automatic annotation of sign language are presented. For gesture segmentation in general, a very complete framework is developed in (Lin et al., 2016). It provides a general overview and a comparison of several works in human motion segmentation using different data sources (camera, MoCap, sensors, etc.) but does not address the problem of ”per channel segmentation” or the particular application of sign language processing.

Continuous signing segmentation and recognition can be opposed to isolated sign language recognition. Coarticulation effects present in continuous signing and absent from isolated signs make the segmentation of the former harder. Most of the existing work on the automatic segmentation of continuous signing relies on video footage to segment at a gloss-level. (Kim et al., 2002) take advantage of Hidden Markov Model (HMM) to segment a continuous stream in Korean Sign Language into signs segments. Similarly, (Yang and Sarkar, 2006) perform sign segmentation of continuous American Sign Language using Conditional Random Fields (CRF). In their article they demonstrate the superiority of the CRF approach (85% accuracy) compared to HMM (60%). A different approach was developed by (Lefebvre-Albaret et al., 2008). It presents a computer-aided segmentation of sign language sequences based on the detection of motion cues such as symmetry, repetition and hand trajectories templates. The algorithm is helped by the punctual intervention of an operator who has to specify one frame belonging to each sign.

However, those segmentation approaches do not take into account the multichannel aspect of sign languages and lead to segmentation schemes highly dependent on the context of the utterance, i.e. the segments implicitly contain the coarticulation effects of the sequential signs. The resulting segmentation is thus hardly reusable in a different context, for example to synthesize new utterances. A lower-level segmentation, based in particular on phonological elements would facilitate sign composition in various contexts in order to produce new utterances. However, although several studies address the issue of the annotation of sign language video data at a gloss level, little attention has been given to the automatic annotation of the different linguistic channels of sign languages. The work of (Stokoe, 1960) gives a phonological structure to signs by specifying three linguistic parameters to describe all signs: hand motion, hand configuration and hand placement. Each feature can take a discrete value in a limited vocabulary. Two complementary features were later added, hand orientation (Battison, 1978) and non manual features such as shoulder tilt or facial expressions. Many phonological structures use those five features to define signs which can be used as a basis for video annotation : the segments are of a finer level than the gloss segmentation and retain a linguistic value. In an early work, (Vogler and Metaxas, 2001) break signs into ”phonemes” (close to the previous five features) and use HMM on the combination of the phonemes to recognize signs. The channels are processed separately but the ultimate purpose is gloss recognition and not channel annotation. In addition, this work is based on the Movement-Hold model which has been later replaced by a more precise phonetic model (Johnson and Liddell, 2011). Furthermore, due to the difficulty of capturing the finger movements, the hand configuration channel was not processed and the authors of the article also chose not to deal with non manual features. More recently, (Dilsizian et al., 2014) propose to add some linguistic knowledge about the composition of lexical signs to considerably improve the performances of the recognition system.

Work on sign language MoCap data is scarcer than on video. For gloss-level segmentation, (Naert et al., 2017) use some kinematic properties of the two wrists that can be extracted from MoCap data. At a finer level, (Héloir et al., 2005) focused on the segmentation of hand configurations using Principal Composant Analysis (PCA) but the work is restricted to fingerspelling segmentation.

The recognition of facial expressions has received increasing attention in recent years, mainly from the computer vision community. Regarding the existing datasets, the availabiliy of 2D recording devices made possible the creation of large data collections, such as the Cohn-Kanade Dataset (posed facial expressions) and its extension (Lucey et al., 2010) (posed + non-posed facial expressions) or the MUG database (Aifanti et al., 2010) (posed + non posed) with many (up to hundreds) actors. High resolution 3D facial databases with expressions have also been created using binocular/structured light cameras (Zhang et al., 2014), (Yin et al., 2008). The frame rate of such optical device is often limited to 60 or less frames per second (usually 20/30 fps) which may be insufficient for those who are interested in dynamic expressive variations. MoCap techniques can capture movements up to 200 fps and more, which makes them much more powerful for analyzing fine expressive variations. Nevertheless, the publicly available facial MoCap databases are still scarce. The multimodal database described in (Busso et al., 2008), which includes facial MoCap with speech and elicited emotional expressions is one of the few existing ones.

This paper presents the automatic annotation of continuous signing in French Sign Language using MoCap data on two linguistic channels : the hand configurations and the facial expressions.

3. Input Data

This section describes the definition and recording of sign language data as well as the specification of the two corpora that have been used for the studies.

3.1. General Considerations

To develop models for automatic annotation, the first solution that might come to mind is to record all the existing signs in all the possible contexts in order to cover exhaustively all the possible cases of sign production. While this can be attempted (with varying degrees of success) for oral languages by, for example, retrieving huge databases from the Internet (e.g. Wikipedia pages, Twitter posts, etc.), this is impossible for sign languages. Indeed, (i) sign language databases are scarce, especially MoCap databases, (ii) sign languages use the 3D space and the temporal dimension which leads to the production of an infinite number of combinations of the different physical channels, and (iii) sign language cannot be limited to their standard, reference signs: many sign language mechanisms such as classifier-predicates are as important as standard signs and depend strongly on the context of the sentence.

Instead of collecting a large set of random data, it might be more relevant to design a corpus specifically suited to the studied problem. One way to reduce the complexity of the capture is to consider each chosen channel separately. Indeed, each channel can display a limited number of different behaviors. For example, we can enumerate a limited number of different hand configurations in sign languages (often less than fifty in French Sign Language). As a consequence, a corpus designed to study and automatically annotate the hand configurations would focus only on a small number of signs to cover all the possible hand configurations.

To sum up, for the application of automatic annotation, a corpus containing many repetitions of a limited number of different occurrences of the studied element will be preferred. The variability induced by a different context in the element production (for example, by capturing the same hand configuration in different signs) or by a different signer, must be recorded in order to improve the generalization capacity of the resulting model.

3.2. Corpora

Two different corpora were used for the presented work: *Sign3D* (Gibet et al., 2015) to annotate hand configurations, and *FEeL*, a novel corpus that is still under development, for the facial expressions.

Characteristics

The *Sign3D* corpus contains eight sequences of motion. Each of these sequences is composed of one to five French Sign Language utterances. The utterances are messages about the opening hours and entrance fees of various town places (swimming pool, museum, etc.), as well as the description of various events (exhibitions, theatre play etc.). The capture was performed on one signer using a *Vicon* MoCap system and an eye-tracking device to follow gaze direction. Facial expressions, body and finger motions were simultaneously recorded during approximately 9 minutes at 100 fps (around 54000 frames in total).

The *FEeL* corpus has been captured using two signers (learner level). Three kinds of sequences - corresponding to different sets of instructions given to the signers - were recorded. We worked exclusively on the affect channel of the face and chose to analyze this channel within the Ekman framework with six categorical classes of basic emotions (i.e. *anger* - *A*, *disgust* - *D*, *fear* - *F*, *joy* - *J*, *sadness* - *Sa*, *surprise* - *Su* and *neutral* - *N*), which was easier to annotate and more understandable by humans than continuous models (e.g. the *Pleasure* - *Arousal* - *Dominance* framework). Three kinds of sequences were recorded:

i) *Isolated Expressions - IE:* the signers were asked to perform a given expression five times, each expression had to be maintained several seconds before returning to neutral (e.g. for joy we have: $N-J-N-J-N-J-N-J-N-J-N$). Six sequences were recorded per signer, one for each class of affect.

ii) *Sequences of Expressions - SE:* the signers were asked to alternate a given expression with each of the five other expressions (e.g. for joy: $N-J-Su-J-A-J-F-J-Sa-J-D-J-N$). Five sequences were recorded per signer.

iii) *Expressive Utterances - EU:* Sign language sentences with emotional content were prepared. The signers were asked to repeat three times each sentence with a given affect (e.g. it was asked to the signer to sign the following sentence with disgust : "There is a spider on my pizza! Yuk!"). 18 sequences were recorded per signer, one for each sentence.

The corpus has been recorded via a *Qualysis* MoCap system. A total of 40 facial markers were tracked at 200 fps.

Manual Annotation

Manual annotations are used as reference and training data for our automatic annotation. It is thus necessary to have a thorough annotation. The *Sign3D* and the *FEeL* corpora have been annotated using the ELAN software (Max Planck Institute for Psycholinguistics, 2017).

The *Sign3D* corpus has been annotated on several channels including, but not restricted to, gloss, hand placement, hand orientation, mouthing, facial expressions and hand configurations. To reduce the error rate and to have a more consistent annotation, two annotators knowledgeable in French Sign Language validated each others' work.

Concerning the annotations of the *FEeL* corpus, we focused our efforts on the affect channel and, so far, a single annotator has been involved in the process. This annotator has been instructed to "subjectively annotate what he saw" with respect to the two following rules: (i) we distinguish two kinds of segments: the *transition* segments where the class vary from a starting expression to an ending expression, and the stable segments where the class doesn't vary along time; (ii) the name of a *transition* segment is the concatenation of the name of the starting class and of the name of the ending class (e.g. *NA* means that the transition come from the *neutral* class to the *anger* class). A stable segment is named according to the maintained expression displayed (e.g. *Sa* stands for *sadness*).

4. Automatic Annotation

This section describes the principal steps to automatically annotate a sign language channel for a given corpus. Motion descriptors are first computed in order to segment and then label the sign language data. The examples of the annotation of the hand configuration and of the affect channels are detailed.

4.1. Descriptors

The raw data collected from motion capture is the vector of the $3D$ positions of the body markers along time and might not be the best representation to study either the hand configurations or the facial expressions. Indeed, it is often required to transform the initial data in order to get a descriptor that depends only on the phenomenon that we intend to analyze. For instance, if the system is supposed to recognize hand configurations, it should not be sensible to morphological differences between the signers.

Hand Configuration Descriptors
While the positions and orientations of the joints vary according to the chosen reference frame, the Euclidean distances between two articulations are invariant. The hand configurations are therefore described by the vector of the Euclidean distances between each joints of each hand. In our model, each hand has 26 joints (five per finger and one for the wrist) resulting in a total of 325 possible combinations. However, some distances are more relevant than others. For example, distances between two consecutive joints (e.g. the second and third joints of the middle finger) are physiologically similar to bones. Those distances only undergo small variations (due to noise in the data) and are not relevant to discriminate hand configurations.
Different subsets of the total number of distances have been tested in order to find the optimal features. A subset of the 29 most discriminating features was preferred (see. fig. 2). It consists of the distances between:

(1) the wrist and the extremities of the fingers (5 distances) to evaluate the bending of the fingers on the palm,

(2) the extremity of one finger with its neighbors (5 distances) to measure the gap between the fingers,

(3) the extremity of each finger and its corresponding knuckle (5 distances) to evaluate the bending of the fingers with respect to the knuckles, and

(4) the extremity of the thumb and the joints of the other fingers (14 distances) to specify the behaviour of the thumb.

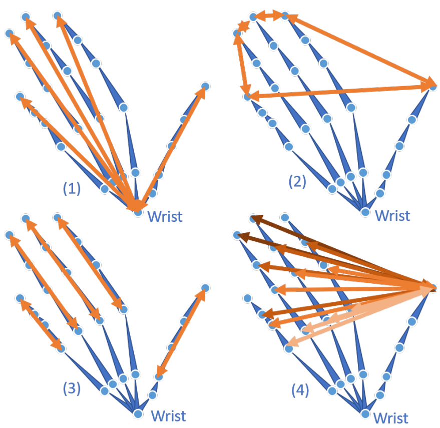


Figure 2: The subset of the 29 distances.

The *Sign3D* corpus that has been used to study the hand configurations contains the data of a unique signer but, in order to have more generic results and to give each distance the same weight, it is necessary to normalize our features. The normalization was performed by dividing each of the 29 types of distances by its maximal value in the corpus. All the distances have therefore a value between 0 and 1. Those distances are then used to segment and label the hand configurations.

Facial Descriptors
A common approach to animate facial expressions of a virtual character is the blendshapes method. An expression $Expr$ can be expressed as the sum of the mesh B_0 representing the neutral expression and a weighted linear combination of n basic deformations b_i expressed differentially from the neutral expression (see also fig. 3):

$$Expr = B_0 + \sum_{i=1}^{n} w_i \cdot b_i \qquad (1)$$

This method has the advantage of providing a light representation (in our case only 51 basic deformations) which leads to faster computations and facilitates storing in our database. In order to automatically obtain the appropriate set of parameters $\{w_{1..n}\}$ at each frame we have to face two problems: i) the targeted avatar and the signer don't have exactly the same morphology (the *retargeting* problem), ii) for one given expression E there might be multiple existing linear combinations $\sum_{i=1}^{n} w_i \cdot b_i$ that minimize the distances between the markers and the corresponding vertices of the mesh (the *non unicity* problem).

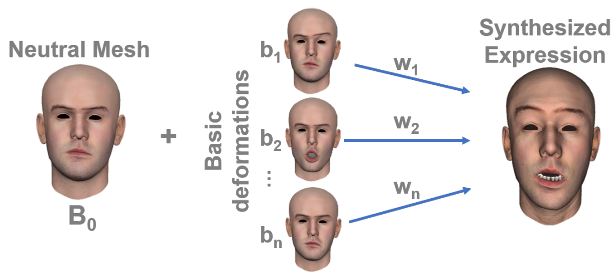


Figure 3: Synthesizing expressions from a linear combination of blendshapes.

We dealt with the *retargeting* problem as in (Bickel et al., 2007) or (Deng et al., 2006). Given one frame where the signer shows a neutral expression, a RBF regression is trained in order to make the link between the position of any point of the signer's face and the position it would have on the avatar's face:

$$\hat{M} = F_{RBF}(M) = \sum_{k=1}^{K} u_k f_k(M) \qquad (2)$$

where $\hat{M}$ is the estimated position of the signer's corresponding marker M retargeted on the avatar's mesh and $\{u_{1..K}\}$ are the optimized weights associated to the radial basis functions $\{f_{1..K}\}$.
The *non unicity* problem is formulated as a minimization problem in which the parameters $\{w_{1..n}\}$ are optimized so that the distances between the retargeted marker positions $\hat{M}_{1..40}$ and the corresponding vertices of the mesh $V_{1..40}$

are reduced. To ensure that the optimal weights found with this method do not generate visual artifacts, some constraints (e.g. non-negativity constraint) and/or some regularization energy that penalizes weights outside the $[0,1]$ range can be incorporated. In our case, we introduced the *Thin-shell* model (Botsch and Sorkine, 2008) as a regularization energy that doesn't directly ensure that the weights stay between 0 and 1 but penalizes the bending and stretching deformations of the initial mesh:

$$\hat{W} = \underset{\{w_{1..K}\}}{\arg\min}(dist_{eucl}(\hat{M}_{1..40}, V_{1..40})^2 + E_{TS}) \quad (3)$$

with $\hat{W}$ the optimal vector of weights $\{w_{1..K}\}$ for the given expression and E_{TS} the *Thin-shell* energy. The vector of blendshape weights $\hat{W}$ is the chosen descriptor for the analysis of the affect channel.

4.2. Automatic Segmentation

The localization of segments of interest in a stream of sign language data is called the *segmentation*. It is done by detecting manually or automatically the temporal points corresponding to the beginning and the end of a behaviour (hand configuration or facial expression in our case). The coarseness of the behaviour to detect depends on the chosen annotation scheme. For example, sign language data can be segmented at a gloss level by detecting the beginning and end of a sign, or at a finer level such as facial expression, by detecting the beginning and end of an affect.

Segmentation of the Hand Configuration Channels

While signing, the signer alternates between stable periods where hand configurations stay the same (no or little motion of the fingers) and transitions between two consecutive hand configurations. The segmentation step therefore consists in separating the continuous signing sequences in two types of segments : *hand configuration* or *transition*. Only the *hand configuration* segments are labeled in the labeling step. To perform the segmentation, we assume that the variation of the distances is discriminating, i.e. the variation will be small during stable configurations and high during transitions. For each frame f, and for each selected distance SD, the variation of the distances $d(i,j)$ between two joints (i and j) is computed between the frame $f-1$ and f. Those variations are summed (see Equation 4 for the right hand RH).

$$VarDist_{f,RH} = \sum_{i \in RH} \sum_{j \in RH d(i,j) \in SD} |d(i,j)_f - d(i,j)_{f-1}| \quad (4)$$

The segmentation relies on the use of a threshold. If *VarDist* is above this threshold, a *transition* segment will be detected. If *VarDist* is below the threshold, the segment will be recognized as a *hand configuration* segment. To select the value of the threshold and to evaluate our segmentation, we used the *Simple Matching Coefficient* (SMC) metric. It measures the similarity between two sets of values (here, the manual and the automatic segmentations). The SMC is the ratio of the number of overlapping frames between the two segmentations on the total number of frames. Figure 4 shows the variation of the SMC of the whole corpus with respect to the chosen threshold for the 29 normalized distances. The maximum (SMC = 81%) is reached for a threshold of 0.2. As manual annotation is performed by a human being on video footage, automatic annotation may be more accurate than manual annotation. Therefore, we consider this threshold satisfactory and it will be the one used in the following steps.

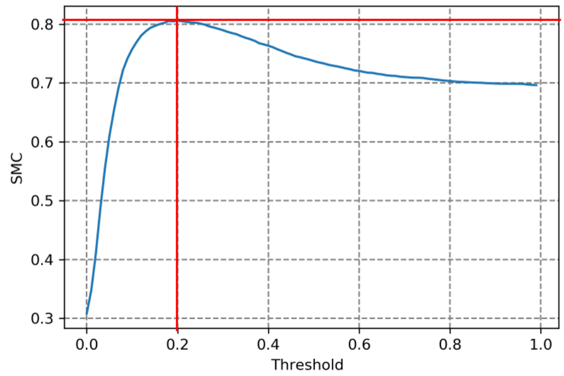


Figure 4: SMC with respect to the threshold values. The maximum (SMC = 81%) is reached for a threshold of 0.2.

Segmentation of the Affect Channel

The affect channel is segmented in a similar way, before the labeling step . We aim at detecting the frames that are located at the border between two segments. Since the border frames are related to the transitions from one expression to another we consider the energy curves of the velocity and acceleration of the n blendshape coefficients:

$$E_{VelBS} = \sum_{i=1}^{n} |\frac{dw_i}{dt}| \quad (5)$$

$$E_{AccBS} = \sum_{i=1}^{n} |\frac{d^2 w_i}{dt^2}| \quad (6)$$

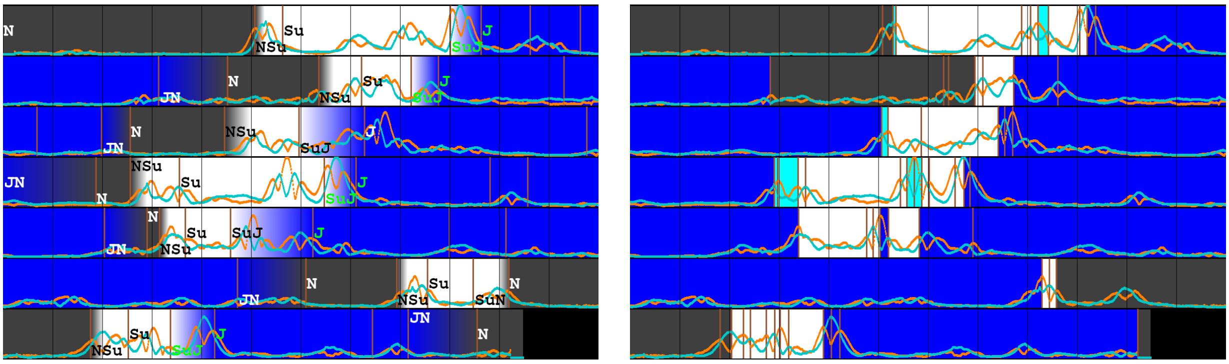
In order to detect the local peaks, we consider the local optima of the curve $E^2_{Vel_{BS}} + E^2_{Acc_{BS}}$, and only keep those for which the local variation is important. This detection procedure is achieved by computing the derivative values on a window surrounding the detected optima, and applying a threshold. The orange and turquoise curves in Figure 5 show an example of segmentation using this method.

4.3. Automatic Labeling

The identification of the previously defined segments of interest is called the *labeling*. This task is highly dependent on the chosen annotation scheme. Typically, it will consists in selecting the right label from a closed vocabulary to identify a segment.

Labeling the Hand Configuration Channels

A supervised machine learning approach is used to identify the handshape on each frame of the *hand configuration* segments. 32 classes were defined corresponding to 32 different handshapes (see fig. 6).



Manually segmented and labeled sequence   Automatically segmented and labeled sequence

Figure 5: An example of automatic annotation: "What? I won 1000 €!" repeated 3 times (white: surprise, blue: joy, cyan: fear; the orange curve stands for the sum of accelerations, the turquoise one for the sum of velocities; the vertical brown lines represent the limits of each segment; each vertical black line stands for 0.5 second).

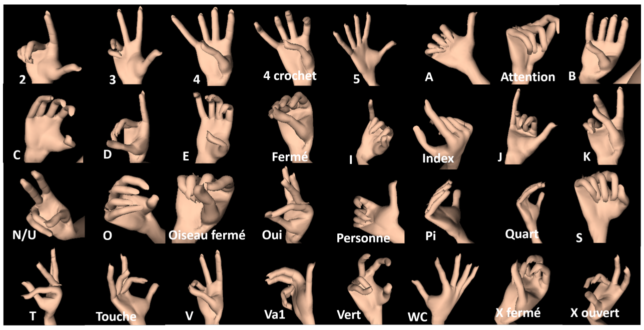


Figure 6: The 32 hand configurations played on an avatar.

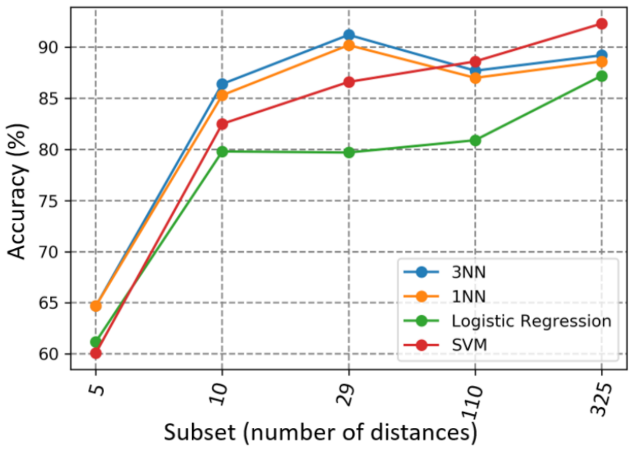


Figure 7: Accuracy on the test set for the hand configurations channel depending on the machine learning algorithm and selected distances.

Like for the segmentation step, the algorithm takes as input the chosen distances between the joints to characterize the hand configurations. Our machine learning classifiers are trained on 23533 manually annotated frames presenting those 32 configurations. The test set is composed of 3927 frames which amount to 14% of the total number of available examples. Our labeling approach sorts each frame in one of the 32 categories. We tested three different machine learning algorithms : logistic regression, support vector machine (SVM) with a *linear* kernel and k-nearest neighbors (kNN) on different subsets of our descriptors. Figure 7 shows the accuracy (number of correct predictions divided by the total number of predictions) on the test set depending on the machine learning algorithm and the subset of distances selected. We can see that the "29 distances" subset presented in Section 4.1. gives the best results with the 3NN approach (91.2%) while the SVM on the 325 distances have the overall best accuracy (92.3%) (but the duration of the classifier training is longer).

Some configurations are more sensible to confusion than others. For example, the 'K' and 'V' configurations are often mistaken (in the two configurations, the middle and index fingers are raised; in the 'K', there is a contact between the thumb and the base of the middle finger while there is not in the 'V'). The 'B' and 'Pi' configurations are also confused as only the thumb position is discriminant between the two configurations.

Labeling the Affect Channel

The facial channel labeling is a supervised learning task aiming at identifying the correct class among the 7 defined in section 3.2.. Different methods were tested: kNN (*1NN* and *3NN*), SVM (*linear* and *RBF* kernels) and Random Forests (RF). The sequences recorded on each signer were processed separately. For each signer, the sequences IE and SE which represent roughly 50% of the data were used as the training set while the EU sequences were used as the test set. Whereas during the training phase, each frame of the training set with its corresponding manually annotated label was considered as a training sample, the test examples were constituted of the average along time of the frames composing each segment. These segments have been previously obtained according to the method presented in section 4.2.. Each of these segments was classified as a whole represented by its average vector of blendshape weights:

$$\bar{\hat{W}} = \frac{\sum_{f=1}^{F} \hat{W}_f}{F} \quad (7)$$

with F the number of *frames* of the considered segment. Figure 5 shows an example of classification using this method; results are detailed in next section.

5. Results

The results of the automatic annotation of the hand configurations and of the facial expression channels are presented in this section.

5.1. Automatic Annotation of the Hand Configurations

To automatically annotate hand configurations during continuous signing, (i) the stream is segmented to distinguish *hand configuration* from *transitions* segments (section 4.2.), then, (ii) the hand configuration of each frame of the *hand configuration* segments is labeled (section 4.3.), and finally, (iii) each *hand configuration* segment is labeled according to the predominant class in the segment (see fig. 8).

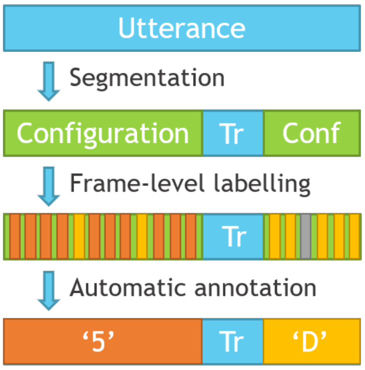


Figure 8: Overview of the automatic annotation of hand configurations.

Figure 9 shows three utterances of French Sign Language manually and automatically annotated. The segmentation threshold has been fixed to 0.2 and the machine learning algorithm used here is 3NN (*3-nearest neighbors*). We worked with the 29 distances described in section 4.1.. While the results can differ from one utterance to another, the recognized labels and segments are mainly consistent with the manual annotation. The results given by the metrics (i.e. accuracy, recall and precision) are computed with respect to the manual annotation and are therefore limited by the 80% overlap of the automatic segmentation with the manual segmentation. There are very few errors in terms of recognition of hand configurations (accuracy of 90%). A perceptual evaluation could give a better assessment of our results.

5.2. Automatic Annotation of the Affect channel

The sequences are first segmented according to the methods described in section 4.2.. Each segment is then labeled independently of the others, according to the following procedure. For each segment, we compute the average vector descriptor over time: $\bar{\vec{W}}$. This vector is used as input of the classification model (kNN or SVM or Random Forest) previously trained on the basis of the learning set. The classifier then returns the label associated with this segment. In order to evaluate the error due to the segmentation, the automatic annotation is performed on both the automatically

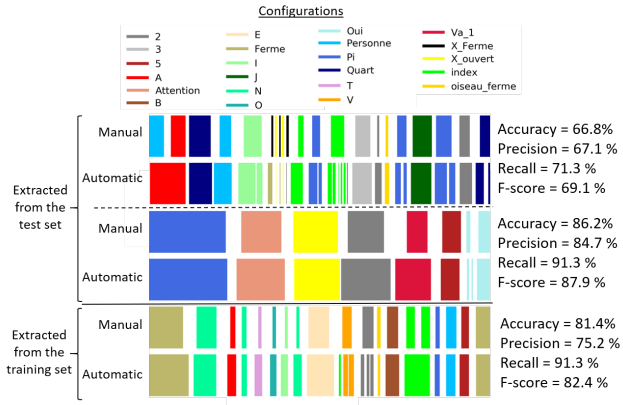


Figure 9: Result of the automatic annotation of hand configurations on three different utterances.

detected segments and the manually defined ones. For both segmentations, Figure 10 gives the accuracy of the classifier for each tested algorithm. It shows that the best results are obtained with the Random Forest algorithm.

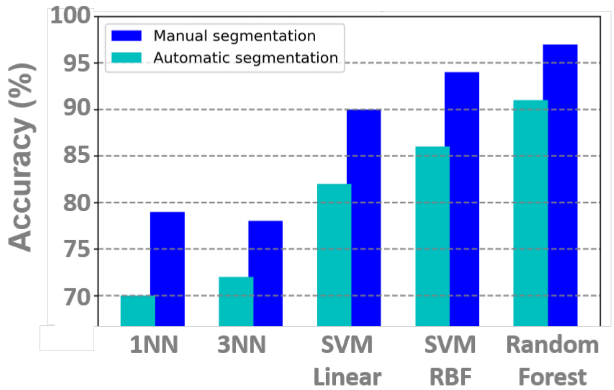


Figure 10: Accuracy on the test set for the affect channel depending on the machine learning algorithms and the segmentation.

6. Conclusion

We designed an approach to automatically annotate sign language MoCap data by processing each channel separately. We detailed the specific examples of hand configuration and facial expression annotation.

There are still many challenges to overcome. Using machine learning models, the automatic annotation could be significantly improved by increasing the size of the dataset, so that the training phase would be more efficient. In addition, following the approaches developed in language processing, we could also use models that learn the dynamics of the sequences, such as Hidden Markov Models, Conditional Random Fields, or Recurrent Neural Networks. However, these methods require large databases.

Another challenge concerns the evaluation of the annotation results. Indeed, for the manual annotation, we rely on a ground truth which may be subject to errors or imprecision. This problem occurs for most recognition or annotation tasks. One solution could be to define a ground truth from a set of previously trained annotators, following strict instructions. In the near future, we plan to validate our annotations by defining quantitative metrics, both for hand configurations and facial expressions. As a complement to assess the quality of the annotation, we also plan to perceptually evaluate the results.

7. Acknowledgements

Part of the observations of this paper are based on the motion capture data and annotations of the Sign3D project (Gibet et al., 2015).

8. References

Aifanti, N., Papachristou, C., and Delopoulos, A. (2010). The mug facial expression database. In 11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10, pages 1–4, April.

Battison, R. (1978). Lexical borrowing in American sign language. ERIC.

Bickel, B., Botsch, M., Angst, R., Matusik, W., Otaduy, M., Pfister, H., and Gross, M. (2007). Multi-scale capture of facial geometry and motion. In ACM SIGGRAPH 2007 Papers, New York, NY, USA.

Botsch, M. and Sorkine, O. (2008). On linear variational surface deformation methods. IEEE Trans. on Visualization and Computer Graphics, 14(1):213–230, January.

Busso, C., Bulut, M., Lee, C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J., Lee, S., and Narayanan, S. (2008). Iemocap: Interactive emotional dyadic motion capture database. Language Resources and Evaluation, 42(4):335–359, 12.

Deng, Z., Chiang, P.-Y., Fox, P., and Neumann, U. (2006). Animating blendshape faces by cross-mapping motion capture data. In Proce. of the 2006 Symposium on Interactive 3D Graphics and Games, pages 43–48, New York, NY, USA.

Dilsizian, M., Yanovich, P., Wang, S., Neidle, C., and Metaxas, D. (2014). A new framework for sign language recognition based on 3d handshape identification and linguistic modeling. In Nicoletta Calzolari (Conference Chair), et al., editors, Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, may. European Language Resources Association (ELRA).

Dreuw, P. and Ney, H. (2008). Towards automatic sign language annotation for the elan tool. In 3rd Workshop on the Representation and Processing of Sign Languages.

Gibet, S., Lefebvre-Albaret, F., Hamon, L., Brun, R., and Turki, A. (2015). Interactive editing in French sign language dedicated to virtual signers: requirements and challenges. Universal Access in the Information Society, 15(4):525–539.

Héloir, A., Gibet, S., Multon, F., and Courty, N. (2005). Captured motion data processing for real time synthesis of sign language. In Motion in Games.

Johnson, R. E. and Liddell, S. K. (2011). A segmental framework for representing signs phonetically. Sign Language Studies, 11(3):408–463.

Kim, J.-B., Park, K.-H., Bang, W.-C., and Bien, Z. (2002). Continuous korean sign language recognition using gesture segmentation and hidden markov model. In Proceedings of the 2002 IEEE International Conference on Fuzzy Systems.

Lefebvre-Albaret, F., Dalle, P., and Gianni, F. (2008). Toward a computer-aided sign segmentation. In Language Resources and Evaluation Conference (LREC). European Language Resources Association.

Lin, J. F. S., Karg, M., and Kulić, D. (2016). Movement primitive segmentation for human motion modeling: A framework for analysis. IEEE Transactions on Human-Machine Systems.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, pages 94–101, June.

Max Planck Institute for Psycholinguistics. (2017). Elan v.4.9.4. `http://tla.mpi.nl/tools/tla-tools/elan/`.

Naert, L., Larboulette, C., and Gibet, S. (2017). Coarticulation analysis for sign language synthesis. In International Conference on Universal Access in Human-Computer Interaction.

Stokoe, W. C. (1960). Sign language structure: An outline of the visual communication systems of the american deaf. Studies in Linguistics, Occasional Papers, 8.

Vogler, C. and Metaxas, D. (2001). A framework for recognizing the simultaneous aspects of american sign language. Computer Vision and Image Understanding, 81(3):358–384.

Yang, R. and Sarkar, S. (2006). Detecting coarticulation in sign language using conditional random fields. In Proceedings of the 18th International Conference on Pattern Recognition.

Yin, L., Chen, X., Sun, Y., Worm, T., and Reale, M. (2008). A high-resolution 3d dynamic facial expression database. In 2008 8th IEEE International Conference on Automatic Face Gesture Recognition, pages 1–6, Sept.

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., Liu, P., and Girard, J. M. (2014). Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. Image and Vision Computing, 32(10):692 – 706. Best of Automatic Face and Gesture Recognition 2013.

NEW Shared & Interconnected ASL Resources: SignStream® 3 Software; DAI 2 for Web Access to Linguistically Annotated Video Corpora; and a Sign Bank

Carol Neidle[1], **Augustine Opoku**[2], **Gregory Dimitriadis**[2], **and Dimitris Metaxas**[2]
[1] Boston University Linguistics Program, [2] Rutgers University Computer Science Department
[1] 621 Commonwealth Ave., Boston, MA 02215; [2]110 Frelinghuysen Rd., Piscataway, NJ 08854-8019
carol@bu.edu, augustine.opoku@gmail.com, gregdimi@cs.rutgers.edu, dnm@cs.rutgers.edu

Abstract

2017 marked the release of a new version of SignStream® software, designed to facilitate linguistic analysis of ASL video. SignStream® provides an intuitive interface for labeling and time-aligning manual and non-manual components of the signing. Version 3 has many new features. For example, it enables representation of morpho-phonological information, including display of handshapes. An expanding ASL video corpus, annotated through use of SignStream®, is shared publicly on the Web. This corpus (video plus annotations) is Web-accessible—browsable, searchable, and downloadable—thanks to a new, improved version of our Data Access Interface: DAI 2. DAI 2 also offers Web access to a brand new Sign Bank, containing about 10,000 examples of about 3,000 distinct signs, as produced by up to 9 different ASL signers. This Sign Bank is also directly accessible from within SignStream®, thereby boosting the efficiency and consistency of annotation; new items can also be added to the Sign Bank. Soon to be integrated into SignStream® 3 and DAI 2 are visualizations of computer-generated analyses of the video: graphical display of eyebrow height, eye aperture, and head position. These resources are publicly available, for linguistic and computational research and for those who use or study ASL.

Keywords: American Sign Language (ASL), linguistically annotated video corpora, annotation software, sign bank

1. Introduction

We report here on several new, interconnected, publicly shared, resources for linguistic and computational analysis of video data from American Sign Language (ASL), developed in conjunction with the American Sign Language Linguistic Research Project (ASLLRP):

- We have released in 2017 a new, improved version of **SignStream®**, the Mac OS software we have been developing for linguistic annotation of ASL video data.[1]
- The annotated corpora are then made available on the Web for viewing, browsing, searching, and downloading via a Web interface that we have developed, our **Data Access Interface (DAI) 2**.[2] The datasets can be downloaded and further analyzed using the SignStream® 3 software that is shared publicly.
- Both SignStream® 3 and DAI 2 now also provide access to a new **ASLLRP Sign Bank**, which makes it possible to view multiple productions, by different ASL signers, of signs of interest. When accessed from within SignStream®, information from the Sign Bank can also be directly entered into the annotations. Furthermore, when new SignStream® datasets are uploaded to DAI 2, the new signs—and new examples of existing signs—are readily added to the Sign Bank.

See the overview in Figure 1.

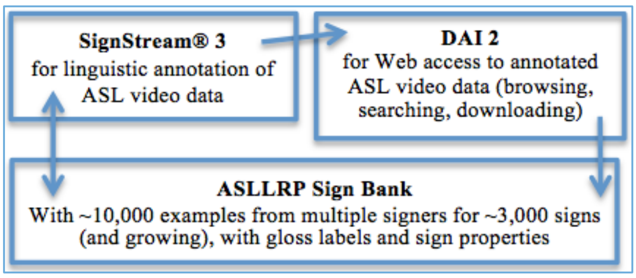


Figure 1. Publicly Shared ASL Linguistic Resources

2. Annotation Software

SignStream® 3 is a Java-based reimplementation of the original Mac Classic software (Neidle, Sclaroff, and Athitsos, 2001; Neidle, 2002), designed for linguistic annotation of video data. SignStream® provides an intuitive interface for labeling and time-aligning manual and non-manual components of the signing. SignStream® 3 has many new features. For example, version 3 enables encoding of morpho-phonological information, including sign type (lexical, fingerspelled, etc.) and number of hands. Handshape information is annotated through use of palettes (specifically for ASL handshapes), and start and end handshapes are displayed as icons left- and right-aligned with the corresponding gloss label; see Figure 2. It is also possible to scroll continuously from one utterance to the next. Version 3 also allows for multiple annotation tiers, well-suited to analysis of dialogs; see Figure 3.

This new Open Source version, released in 2017, is available from http://www.bu.edu/asllrp/SignStream/3/ and requires MacOS 10.8 or higher. For further details about functionalities, see (Neidle, 2017).

3. Interfaces for Web Access to Corpora

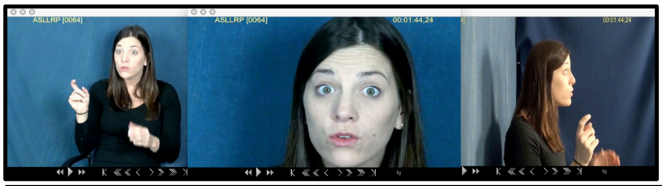
We previously developed a Web-based **Data Access Interface (DAI)**[3] for sharing our ASL video corpora created with SignStream® 2. The DAI facilitates browsing, search, and download of the data (Neidle and Vogler, 2012). The DAI was extended to provide access as well to the **American Sign Language Lexicon Video Dataset (ASLLVD)**, with ~10,000 citation-form examples (of ~3,000 signs) (Neidle, Thangali, and Sclaroff, 2012).

We have recently created a new **Data Access Interface, DAI 2**,[4] because the new version of SignStream® incorporates significant enhancements to the annotations (now including handshape information, e.g.). Thus, the DAI needed to be extended for display of the richer repre-

[1] Gregory Dimitriadis is the principal developer for version 3.
[2] Augustine Opoku is the principal developer for DAI 2.

[3] http://secrets.rutgers.edu/dai/queryPages/
[4] http://dai.cs.rutgers.edu

sentations in our new **ASLLRP SignStream® 3 Corpus**. We have taken the opportunity to provide more powerful search functionalities, as well. It is now possible to search for characters in the gloss string (on the dominant and/or non-dominant hands), and type of sign (e.g., lexical, fingerspelled, classifiers, or specific types of classifier, and to restrict the search to 1- or 2-handed signs and/or signs containing a particular start and/or end handshape on either or both hands. Searches can also be restricted to particular data sources or signers. It is also possible to search for utterances that contain specific types of non-manual events (e.g., raised eyebrows) or grammatical markings (e.g., wh-question). The user can select the view (front, side, close-up of the face) and play the video of the sign or the utterance containing the sign. This is shown in Figure 4.



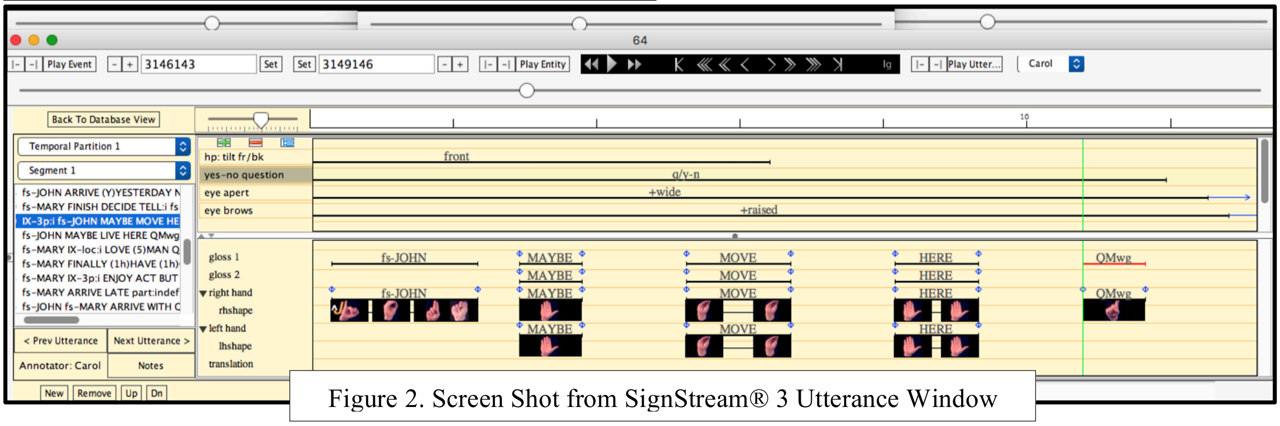


Figure 2. Screen Shot from SignStream® 3 Utterance Window

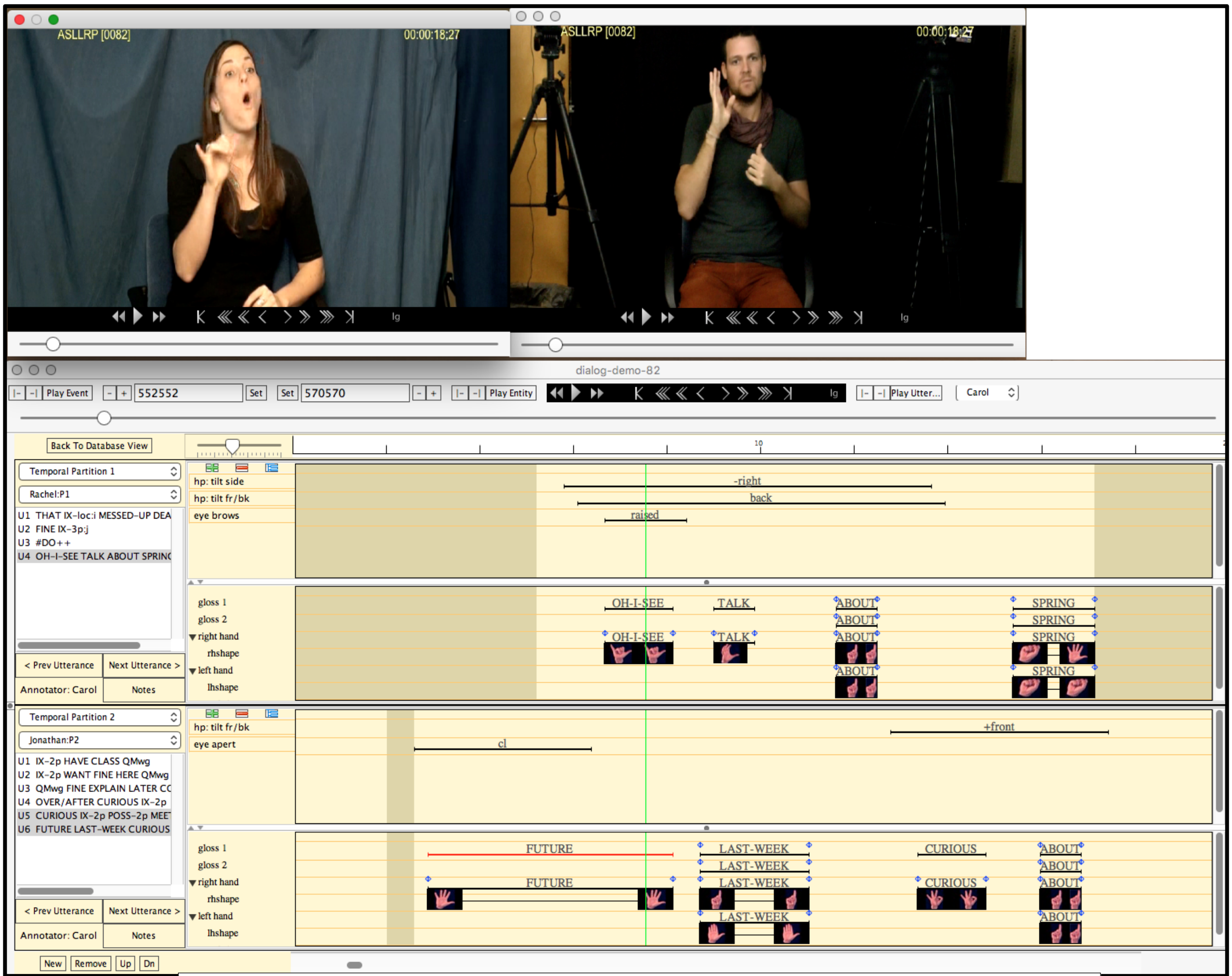


Figure 3. Multiple Tiers – Facilitating Annotation of Dialogs within SignStream®

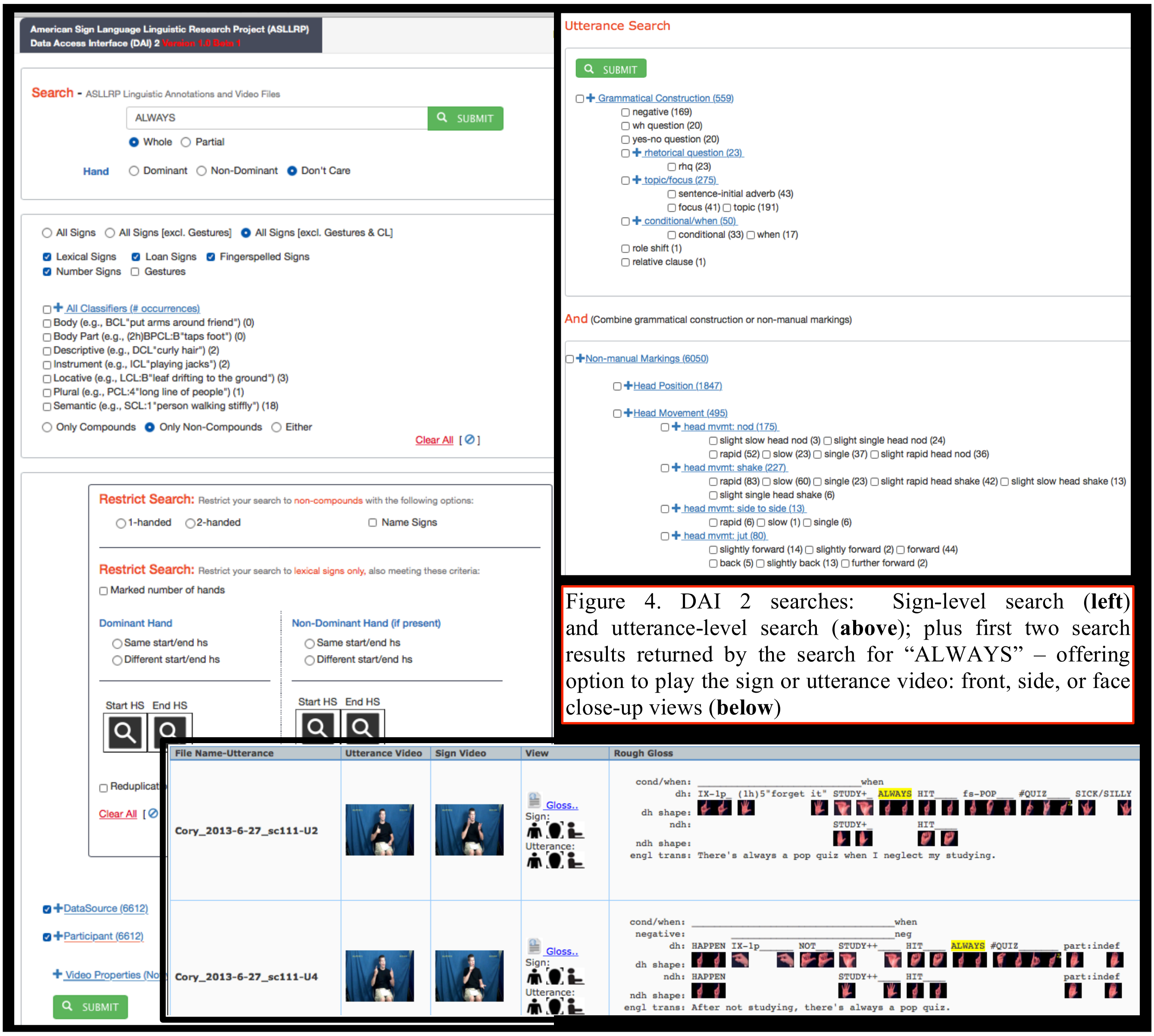


Figure 4. DAI 2 searches: Sign-level search (**left**) and utterance-level search (**above**); plus first two search results returned by the search for "ALWAYS" – offering option to play the sign or utterance video: front, side, or face close-up views (**below**)

Material of interest can be designated for download according to user preferences. The download feature gives the user the ability to add utterances or entire SignStream® files to the download cart while browsing the various search results. The user can then initiate the download of the marked items from the download page (after selecting options, such as choice(s) of video views (Front, Face, Side)). The associated components (including video files and annotation, in XML format) will then be packaged and returned to the user in a compressed format (zip). The user can save the packages and return to them at a later date. This allows users to browse and add items to the cart on a low bandwidth connection and return to download the packaged items when they are on a faster Internet connection. The user can also create a download package and share the link to it with other users who can subsequently view and download the items in that package. After SignStream® files have been downloaded, they can also be opened using the SignStream® software to allow for further exploration by the user.

4. Sign Bank

DAI 2 also provides access to a new ASLLRP Sign Bank. The Sign Bank was initially comprised of the data from the ASLLVD, the American Sign Language Lexicon Video Dataset (Sclaroff et al., 2010; Thangali et al., 2011; Neidle, Thangali, and Sclaroff, 2012), consisting of almost 10,000 examples of almost 3,000 distinct signs, in citation form. However, DAI 2 provides a simple mechanism for adding new signs to the Sign Bank as new data get added to our continuous signing corpora. Since these signs are clipped from continuous video, however, they are different in appearance from recordings of signs produced in citation form. In the future, we plan to video-record citation-form examples of the newer. Nonetheless, for the time being, this allows us to expand the collection of signs and signers in the Sign Bank and also to offer users examples of sentences containing particular signs.

Figure 5 shows a Sign Bank search via DAI 2. The user can search for a text string, and for properties of the sign, including start and end handshapes. The search results are displayed; it is possible to view any or all of the examples of a given sign, as well as the containing utterances.[5]

5. Access to the Sign Bank from within SignStream®

SignStream® users can search the Sign Bank for the sign they wish to annotate. See Figure 6 below. Thus users can ensure that the gloss label chosen is consistent with the glossing of previous examples of the same sign. Furthermore, if the desired sign is found in the Sign Bank, then it can be entered directly into the annotation with its associated properties and handshapes. The user can further edit if modifications are necessary. If the sign in question is not already in the Sign Bank, the user can add the sign to their local Sign Bank so the information will be available for subsequent annotations.

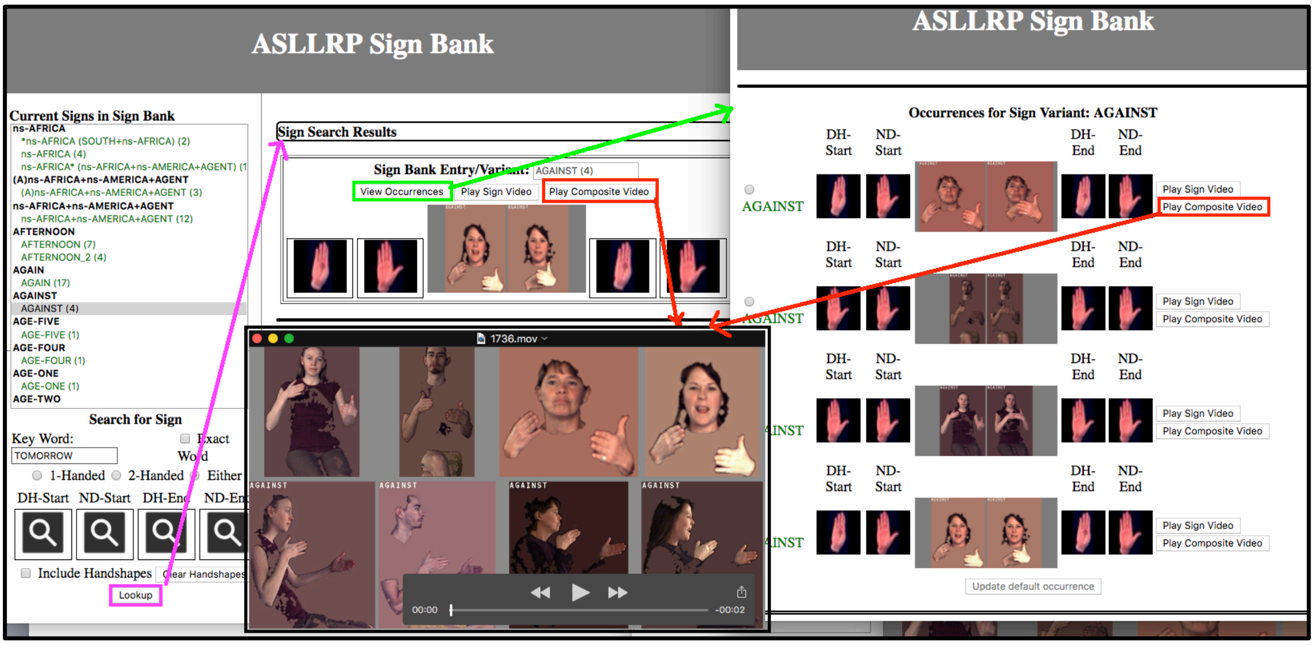


Figure 5. Sign Bank–Access from DAI 2: Sample Search for Text String AGAINST in Gloss. User can display all occurrences and play sign videos or composite video of all productions together.

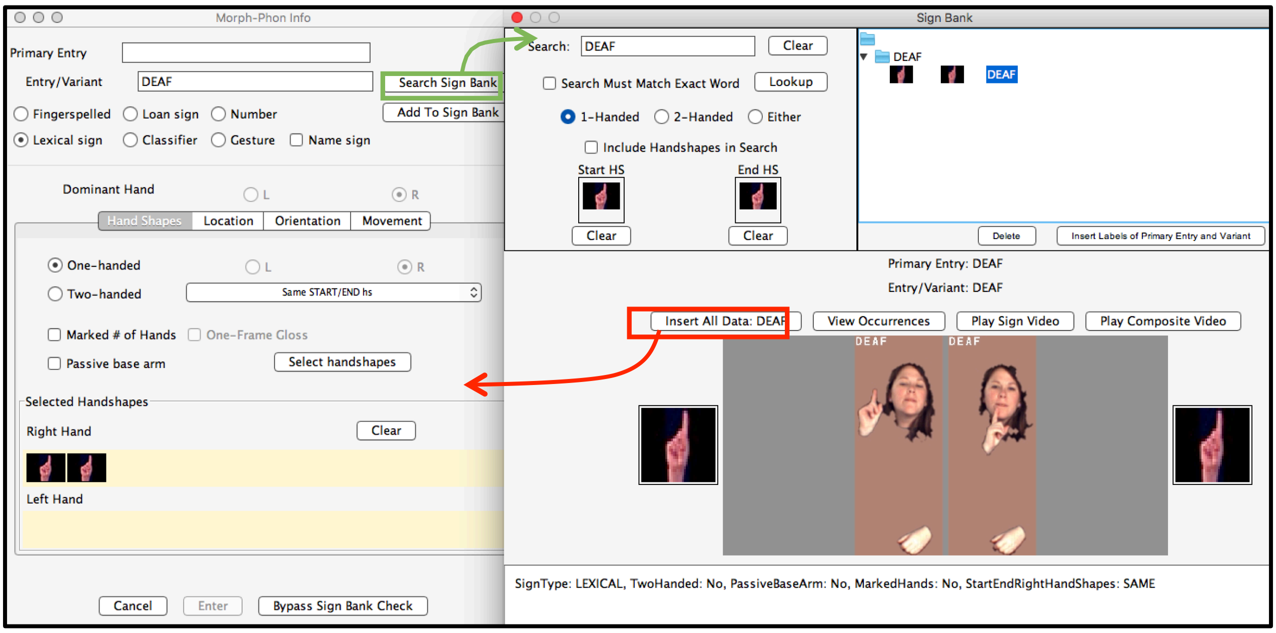


Figure 6. Sign Bank–Access from within SignStream®: Sign labels & properties can be copied directly into annotations

[5] There is another "sign bank" project under development for ASL (Hochgesang, Crasborn, and Lillo-Martin, 2017), but this is not yet shared publicly, so it is difficult to compare with ours. "Sign bank" projects for other signed languages (e.g., Auslan, BSL, NGT, FinSL, and Swiss German Sign Language (DSGS)) are somewhat different in nature from ours; they tend to be more dictionary-like (see, e.g., https://github.com/signbank/).

6. Available New Data

The **ASLLRP SignStream® 3 Corpus** is shared through the DAI 2 interface. It is an expanding collection; files are added as verifications of the annotations are completed. The corpus includes 3 different ASL signers, and the shared data (as of February 2018) include over 6,000 sign tokens, in just over 300 total utterances, from 2 signers.

The data were elicited in an open-ended way. We explained to our ASL consultants that we were interested in a wide range of different types of constructions (e.g., questions, negations, conditional sentences, etc.) and they were asked to come up with a set of sentences that were natural for them to produce. They were given no specific directions about content or structure. Subsequent signers were shown the examples produced by the earlier signers and asked to produce, in general, similar types of sentences.

7. Value for Research, Education, and Potential Future Applications

The video data and annotations have been used by our extended research team and by others for linguistic and computational research on ASL. Linguistic and computer science research by others (including students) that has made use of our data and software over the years includes, e.g., among many others: (Goldenstein, Vogler, and Velho, 2005; Vogler and Goldenstein, 2005; Zahedi et al., 2005; Zahedi, Keysers, and Ney, 2005a; b; Goldenstein and Vogler, 2006; Grossman and Kegl, 2006; Rybach, 2006; Zahedi et al., 2006a; Zahedi et al., 2006b; Ciaramello and Hemami, 2007; Davidson, Caponigro, and Mayberry, 2008; Forster, 2008; Hendriks, 2008; Roh and Lee, 2008; Vogler and Goldenstein, 2008b; a; Weast, 2008; Williford, 2008; Yang, Sclaroff, and Lee, 2009; Yang and Lee, 2010; Caponigro and Davidson, 2011; Kammann, 2012; Nguyen and Ranganath, 2012; Greene, 2013; Yang and Lee, 2013; Wolfe et al., 2014; Roush, 2015; Toman and Kuefler, 2015; Boulares and Jemni, 2016; Costello, 2016; Kim et al., 2016; Lim, Tan, and Tan, 2016b; a; Raud, 2016; Roush, 2016; Elakkiya and Selvamani, 2017; Kumar, 2017).

Our own research on computer-based recognition of manual signs and of non-manual grammatical information has also greatly benefited from use of these data, e.g.: (Athitsos, 2006; Duffy, 2007; Thangali et al., 2011; Liu et al., 2012; Metaxas et al., 2012 ; Liu et al., 2013; Thangali, 2013; Dilsizian et al., 2014; B. Liu et al., 2014; J. Liu et al., 2014; Neidle et al., 2014; Mark Dilsizian et al., 2016; M. Dilsizian et al., 2016; Yanovich, Neidle, and Metaxas, 2016; Metaxas, Dilsizian, and Neidle, 2018). Most recently, we have shown high accuracy and scalability in recognition of signs from our Sign Bank, using model-based machine learning, with incorporation of linguistically relevant features and constraints (Metaxas, Dilsizian, and Neidle, 2018). For a vocabulary of 350 signs from our Sign Bank, we achieve recognition accuracy of 93.3%. In 97.9% of the cases, the correct sign is within the top 5 results.

What this means is that we can envision development of a user interface that would allow a user to search for a sign in our Sign Bank in one of two ways: either by producing the sign in front of a webcam, or by selecting a sign by identifying its start and end points from a continuous video. The user could then be offered 5 (e.g.) likely options, in order of decreasing likelihood, with the option to play any of those signs to confirm or disconfirm the correctness of the sign identification. This is illustrated in Figure 7. The user could then be taken to the relevant information in our Sign Bank. This could also be used from within SignStream® to facilitate the annotation process, especially for signs that the user may not know how to gloss. This interface could also be used as an entryway to other ASL resources, e.g., to enable sign lookup in an ASL dictionary. We intend to pursue research to make such a lookup interface a reality.

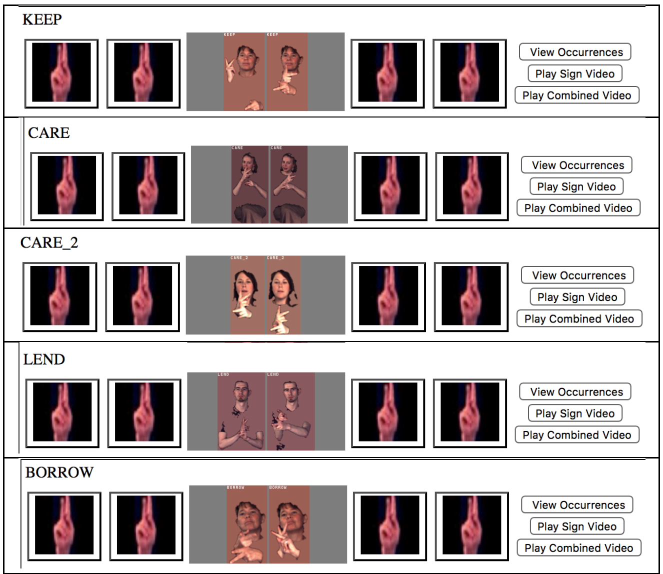


Figure 7. Example of Choices to be Offered to a User before Confirmation of a Lookup Selection – Based on the Interface for our Current ASLLRP Sign Bank

These tools also have obvious applications to education, for those teaching/learning ASL.

8. Planned Enhancements

In addition to developing lookup capabilities just described for navigation through our own resources, we are also currently working to expand the functionalities of both SignStream® and DAI 2 to allow display of computer-generated analyses of the relevant video. In particular, we now have the ability to produce graphs from the close-up face view to illustrate changes, over time, in eyebrow height, eye aperture, and head rotation along the 3 axes. See Figure 8 and our website with examples (ASLLRP, 2016). This will provide valuable information for linguis-

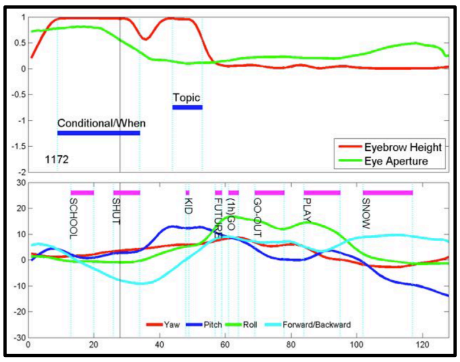


Figure 8. Computer-generated Graphical Information about Facial Expressions

tic and computational research on ASL of a kind that has not been available to date over large datasets Ultimately such technology will also enable semi-automatic transcription of sign language data.

9. Acknowledgements

Special thanks to Christian Vogler (Gallaudet University) for his implementation of the original DAI, and to Stan Sclaroff and Ashwin Thangali for providing support at Boston University and especially for their important roles in the development of the ASLLVD. For a more comprehensive list of the language consultants and the many students and colleagues at Boston, Rutgers, and Gallaudet Universities who have contributed to these efforts, see http://www.bu.edu/asllrp/SignStream/3. This research has been partially supported by the National Science Foundation (grants #1748016, 1748022, 1059218, 0705749, 0958247, 1703883, 1567289, 1555408, 1451292, 1447037).

8. Bibliographical References

ASLLRP (2016) The American Sign Language Linguistic Resource Project, ASL feature tracking and identification of non-manual expressions of linguistic significance : Data visualizations. http://www.bu.edu/av/asllrp/NM/.

Athitsos, V. (2006) Learning Embeddings for Indexing, Retrieval, and Classification, with Applications to Object and Shape Recognition in Image Databases. Doctoral Dissertation, Boston University, Boston, MA.

Boulares, M. and Jemni, M. (2016) Learning sign language machine translation based on elastic net regularization and latent semantic analysis. *Artificial Intelligence Review, 46*(2), pp. 145-166.

Caponigro, I. and Davidson, K. (2011) Ask, and tell as well: Question–Answer Clauses in American Sign Language. *Natural Language Semantics, 19*(4), pp. 323–371.

Ciaramello, F.M. and Hemami, S.S. (2007). 'Can you see me now?' An objective metric for predicting intelligibility of compressed American Sign Language video. *Proceedings of the SPIE Vol. 6492, Human Vision and Electronic Imaging 2007, San Jose, CA*.

Costello, B.D.N. (2016) Language and modality: Effects of the use of space in the agreement system of lengua de signos española (Spanish Sign Language). Unpublished PhD, University of Amsterdam.

Davidson, K., Caponigro, I., and Mayberry, R. (2008). Clausal question-answer pairs: Evidence from American Sign Language. *Proceedings of the 27th West Coast Conference on Formal Linguistics (WCCFL 27)*.

Dilsizian, M., Yanovich, P., Wang, S., Neidle, C., and Metaxas, D. (2014). New Framework for Sign Recognition based on 3D Handshape Identification and Linguistic Modeling. *Proceedings of LREC 2014, Reykjavik, Iceland. May 2014.*

Dilsizian, M., Tang, Z., Metaxas, D., Huenerfauth, M., and Neidle, C. (2016). The Importance of 3D Motion Trajectories for Computer-based Sign Recognition. *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining. LREC 2016. Slovenia, May 2016.*

Dilsizian, M., Tang, Z., Metaxas, D., Huenerfauth, M., and Neidle, C. (2016). The Importance of 3D Motion Trajectories for Computer-based Sign Recognition. *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining. LREC 2016, Portorož, Slovenia. May 2016.*

Duffy, Q. (2007) The ASL Perfect Formed by Preverbal FINISH. Unpublished MA thesis, Boston University, Boston, MA.

Elakkiya, R. and Selvamani, K. (2017) Extricating Manual and Non-Manual Features for Subunit Level Medical Sign Modelling in Automatic Sign Language Classification and Recognition. *Journal of Medical Systems, 41*.

Forster, J. (2008) An Integrated Tracking And Recognition Approach For Video. Unpublished Master's Thesis, RWTH Aachen University, Germany.

Goldenstein, S., Vogler, C., and Velho, L. (2005) Adaptive Deformable Models for Graphics and Vision. *Computer Graphics Forum, 24*(4), pp. 729-741.

Goldenstein, S. and Vogler, C. (2006). When Occlusions are Outliers. *Proceedings of the Workshop on the 25 Years of RANSAC (in conjunction with CVPR)*.

Greene, D.J. (2013) Keeping it Vague: A Study of Vague Language in an American Sign Language Corpus and Implications for Interpreting between American Sign Language and English. Unpublished MA thesis, Western Oregon University.

Grossman, R.B. and Kegl, J. (2006) To Capture a Face: A Novel Technique for the Analysis and Quantification of Facial Expressions in American Sign Language. *Sign Language Studies, 6*(3), pp. 273-305.

Hendriks, B. (2008) Jordanian Sign Language: aspects of grammar from a cross-linguistic perspective. Unpublished.

Hochgesang, J.A., Crasborn, O., and Lillo-Martin, D. (2017) Building the ASL Signbank on the Shoulders of Giants. Poster presented at the CL2017 Pre-Conference Workshop 3: Corpus-based approaches to sign language linguistics: Into the second decade. University of Birmingham, July 24.

Kammann, J. (2012) The Role of Phonology and Semantics in the Leixcal Processing of ASL Core lexical Items and Classifier Constructions. Unpublished MA thesis, University of New Mexico.

Kim, J.-H., Kim, N., Park, H., and Park, J.C. (2016) Enhanced Sign Language Transcription System via Hand Tracking and Pose Estimation. *Journal of Computing Science and Engineering, September 2016, 10*(3), pp. 95-101.

Kumar, N. (2017). Motion Trajectory based Human Face and Hands Tracking for Sign Language Recognition. *Proceedings of the the 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON) GLA University, Mathura, Oct 26-28, 2017*.

Lim, K.M., Tan, A.W.C., and Tan, S.C. (2016a) Block-based histogram of optical flow for isolated sign language recognition. *J. Vis. Commun. Image R., 40*, pp. 538–545.

Lim, K.M., Tan, A.W.C., and Tan, S.C. (2016b) A feature covariance matrix with serial particle filter for isolated sign language recognition. *Expert Systems with Applications, 54*, pp. 208–218.

Liu, B., Kosmopoulos, D., Dilsizian, M., Yang, P., Neidle, C., and Metaxas, D. (2012) Detection and Classification of Non-manual Grammatical Markers in American Sign Language (ASL). In, Poster presented at the 2nd Multimedia and Vision Meeting in the Greater New York Area, Columbia University, New York City, NY, June 15, 2012.

Liu, B., Liu, J., Yu, X., Metaxas, D., and Neidle, C. (2014). 3D Face Tracking and Multi-scale, Spatio-temporal Analysis of Linguistically Significant Facial Expressions and Head Positions in ASL. *Proceedings of LREC 2014, Rejkjavik, Iceland. May 2014*.

Liu, J., Liu, B., Zhang, S., Yang, F., Yang, P., Metaxas, D.N., and Neidle, C. (2013) Non-manual Grammatical Marker Recognition based on Multi-scale Spatial Temporal Analysis of Head Pose and Face. In *Proceedings of the 10th IEEE International Conference on Automatic Face and Gesture Recognition, Shanghai, China, April 25, 2013*.

Liu, J., Liu, B., Zhang, S., Yang, F., Yang, P., Metaxas, D.N., and Neidle, C. (2014) Non-manual Grammatical Marker Recognition based on Multi-scale Spatio-temporal Analysis of Head Pose and Facial Expressions. *Image and Vision Computing Journal (Special issue: The Best of Face and Gesture 2013), 32*(10), pp. 671-681.

Metaxas, D., Liu, B., Yang, F., Yang, P., Michael, N., and Neidle, C. (2012). Recognition of Nonmanual Markers in American Sign Language (ASL) Using Non-Parametric Adaptive 2D-3D Face Tracking. *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. LREC 2012, Istanbul, Turkey. May 2012*.

Metaxas, D., Dilsizian, M., and Neidle, C. (2018). Linguistically-driven Framework for Computationally Efficient and Scalable Sign Recognition. *Proceedings of LREC 2018, Miyagawa, Japan. May 2018*.

Neidle, C., Sclaroff, S., and Athitsos, V. (2001) SignStream™: A Tool for Linguistic and Computer Vision Research on Visual-Gestural Language Data. *Behavior Research Methods, Instruments, and Computers, 33*(3), pp. 311-320.

Neidle, C. (2002) SignStream™: A Database Tool for Research on Visual-Gestural Language. *Journal of Sign Language and Linguistics, 4*(1/2), pp. 203-214.

Neidle, C., Thangali, A., and Sclaroff, S. (2012). Challenges in Development of the American Sign Language Lexicon Video Dataset (ASLLVD) Corpus. *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. LREC 2012, Istanbul, Turkey. May 2012*.

Neidle, C. and Vogler, C. (2012). A New Web Interface to Facilitate Access to Corpora: Development of the ASLLRP Data Access Interface (DAI). *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. LREC 2012, Istanbul, Turkey. May 2012.*

Neidle, C., Liu, J., Liu, B., Peng, X., Vogler, C., and Metaxas, D. (2014). Computer-based tracking, analysis, and visualization of linguistically significant non-manual events in American Sign Language (ASL). *Proceedings of the 6th Workshop on the Representation of Signed Languages: Beyond the Manual Channel. LREC 2014, Rejkjavik, Iceland. May 2014.*

Neidle, C. (2017) A User's Guide to SignStream® 3. Boston, MA: American Sign Language Linguistic Research Project Report No. 16, Boston University.

Nguyen, T.D. and Ranganath, S. (2012) Facial expressions in American Sign Language: Tracking and recognition. *Pattern Recognition, 45*(5), pp. 1877–1891.

Raud, A.L. (2016) Spatial Metaphors in American English and American Sign Language. Unpublished, BA Thesis, Tallinn University, Estonia.

Roh, M.-C. and Lee, S.-W. (2008) A Semi-Dynamic Bayesian Network for Human Gesture Recognition. *2008 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2008)*, pp. 644-649.

Roush, D.R. (2015) The Translation of Event-Structure Metaphors Rendered by Deaf Translators from English to American Sign Language. In *Order No. 3663049 Gallaudet University, 2015.*

Roush, D.R. (2016) The Expression of the Location Event-Structure Metaphor in American Sign Language. *Sign Language Studies, 16*(3), pp. 389-432.

Rybach, D. (2006) Appearance-Based Features for Automatic Continuous Sign Language Recognition. Unpublished Master's Thesis, RWTH Aachen University, Germany.

Sclaroff, S., Athitsos, V., Neidle, C., Nash, J., Stefan, A., Thangali, A., Wang, H., and Yuan, Q. (2010) American Sign Language Lexicon Project: Video Corpus and Indexing/Retrieval Algorithms. Poster presented at the International Workshop on Computer Vision (IWCV), Vietri Sul Mare, Salerno, Italy. May 25-27, 2010.

Thangali, A., Nash, J.P., Sclaroff, S., and Neidle, C. (2011). Exploiting Phonological Constraints for Handshape Inference in ASL Video. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) 2011*.

Thangali, A. (2013) Exploiting Phonological Constraints for Handshape Recognition in Sign Language Video. Unpublished, Doctoral Dissertation, Boston University.

Toman, P. and Kuefler, A. (2015) Classifying Non-manual Markers in American Sign Language. In *Class project, Stanford University*, http://cs229.stanford.edu/proj2015/154_report.pdf.

Vogler, C. and Goldenstein, S. (2005) Analysis of facial expressions in American Sign Language. In: Proceedings of the 3rd Intl. Conf. on Universal Access in Human-Computer Interaction (UAHCI) 2005.

Vogler, C. and Goldenstein, S. (2008a) Facial movement analysis in ASL. *Universal Access in the Information Society, 6*(4), pp. 363-374.

Vogler, C. and Goldenstein, S. (2008b) Toward computational understanding of sign language. *Technology and Disability: Electronic speech processing for persons with disabilities, 20*(2), pp. 109-119.

Weast, T.P. (2008) Questions in American Sign Language: A quantitative analysis of raised and lowered eyebrows. Unpublished Doctoral dissertation, University of Texas, Arlington.

Williford, L.L. (2008) Frequency of Classifier Constructions in American Sign Language. Unpublished MA Thesis, University of Pittsburgh, PA.

Wolfe, R., McDonald, J.C., Berke, L., and Stumbo, M. (2014). Expanding n-gram analytics in ELAN and a case study for sign synthesis. *Proceedings of LREC 2014*.

Yang, H.-D., Sclaroff, S., and Lee, S.-W. (2009) Sign Language Spotting with a Threshold Model Based on Conditional Random Fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), July 2009, 31*(7), pp. 1264-1277.

Yang, H.-D. and Lee, S.-W. (2010) Simultaneous spotting of signs and fingerspellings based on hierarchical conditional random fields and boostmap embeddings. *Pattern Recognition, 43*(8), pp. 2858–2870.

Yang, H.-D. and Lee, S.-W. (2013) Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine. *Pattern Recognition Letters, 34*(16), pp. 2051-2056.

Yanovich, P., Neidle, C., and Metaxas, D. (2016). Detection of Major ASL Sign Types in Continuous Signing for ASL Recognition. *Proceedings of LREC 2016, Portorož, Slovenia. May 2016.*

Zahedi, M., Keysers, D., Deselaers, T., and Ney, H. (2005) Combination of Tangent Distance and an Image Distortion Model for Appearance-Based Sign Language Recognition. In *Pattern Recognition*, Berlin / Heidelberg: Springer. pp. 401-408.

Zahedi, M., Keysers, D., and Ney, H. (2005a). Pronunciation clustering and modeling of variability for appearance-based sign language recognition. *Proceedings of the International Gesture Workshop 2005, Vannes, France, May 2005*.

Zahedi, M., Keysers, D., and Ney, H. (2005b). Appearance-based recognition of words in American Sign Language. *Proceedings of the IbPRIA 2005, 2nd Iberian Conference on Pattern Recognition and Image Analysis, June 2005.,* Estoril, Portugal.

Zahedi, M., Dreuw, P., Rybach, D., Deselaers, T., Bungeroth, J., and Ney, H. (2006a). Continuous Sign Language Recognition – Approaches from Speech Recognition and Available Data Resources. *Proceedings of the Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios, LREC 2006, Genoa, Italy*.

Zahedi, M., Dreuw, P., Rybach, D., Deselaers, T., and Ney, H. (2006b). Geometric Features for Improving Continuous Appearance-based Sign Language Recognition. *Proceedings of the British Maschine Vision Conference, Edinburgh, UK*.

Animating AZee Descriptions Using Off-the-Shelf IK Solvers

Fabrizio Nunnari[1], **Michael Filhol**[2], **Alexis Heloir**[1,3]
[1]DFKI/MMCI/SLSI Group, Saarbrücken, Germany,
[2]LIMSI, CNRS, Université Paris–Saclay, Univ. Paris–sud, bât. 508, 91400 Orsay, France,
[3]LAMIH UMR CNRS 8201 Université de Valenciennes, 59300 Le Mont Houy, France
fabrizio.nunnari@dfki.de, michael.filhol@limsi.fr, alexis.heloir@dfki.de

Abstract

We propose to implement a bottom-up animation solution for the AZee system. No low-level AZee animation system exists yet, which hinders its effective implementation as Sign Language avatar input. This bottom-up approach delivers procedurally computed animations and, because of its procedural nature, it is capable of generating the whole possible range of gestures covered by AZee's symbolic description. The goal is not to compete on the ground of naturalness since movements are bound to look robotic like all bottom-up systems, but its purpose could be to be used as the missing low-level fallback for an existing top-down system. The proposed animation system is built on the top of a freely available 3D authoring tool and takes advantage of the tool's default IK solving routines.

Keywords: Sign Language Synthesis, Signing Avatar, AZee.

1. Introduction on Signing Avatars

In signing avatar technology, current approaches for the creation of sign repositories can be generally described as *pre-animated* or *synthesised*. Solutions embracing the pre-animated approach start from an analysis of full sentences, which are then segmented at a coarse, lemma level. Very large repositories are populated by captured or manually authored Sign Language (SL) animation clips. SL generation is then performed by sequentially stitching together the available segments. In contrast, solutions embracing the synthesized approach are derived from a linguistic representation. This leads to a concise set of atomic animation elements which are symbolically described in a high level declarative language. In this case, SL generation consists of a procedural realization of the symbols composing first signs and then full sentences.

Because they have been manually authored or captured on human performers, pre-animated approaches usually deliver animations which are perceived by end-users as more realistic and natural. However, they cannot extrapolate much beyond the set of low level pre-stored animation clips. On the contrary, synthesised approaches deliver procedurally computed animation which are perceived as stiff and robotic by end-users, but, because of their procedural nature, they are presumably capable of generating the whole possible range of gestures covered by their symbolic description.

In both cases, a clear requirement for any system today is that it be able to animate various articulators on the body and face, with flexible timing patterns. By this we mean, for example, be able to control all communicative channels simultaneously (hands, eyes, lips, and others), but *not* all sharing the same time boundaries.

In this paper, after a survey of the related work (Section 2.), we present in Section 3. how the AZee system is designed to: i) realize animations that feature interleaving communicative channels, and ii) is designed to be part of a generator interchanging synthesis with pre-animated sequences. This novel approach has been so far only a design proposal. In Section 4., we describe how we used a popular open-source 3D editor and its integrated Inverse Kinematic solver to generate AZee animations. The implementation is a work-in-progress. Section 5. illustrates preliminary results in the realization of static poses together with performance tests. Finally, Section 6. concludes the paper.

2. Related Work

Signing avatars have been under development for more than a decade. One of the first working systems was JASigning[1] from the Visicast (Jennings et al., 2010) and the DictaSign (Efthimiou et al., 2010) projects: it uses only the *synthesised* approach and is able to produce signing animation from SiGML (Hanke, 2004) statements as input. SiGML is a digital representation of the Hamburg Notation System (HamNoSys) (Prillwitz et al., 1989), which is a graphical formalism for the description of Sign Language using a set of pictograms. By design, HamNoSys is an oversimplification of sign language, describing sentences as a sequence of glosses. Only within glosses there is a parallelization between manual (hands, fingers) and non-manual features (eyes, lips, nodding, ...). As such, resulting animations are generally perceived as unnatural.

The following technologies tried to overcome the strong limitations of approaches based on pure-synthesis (i.e., the complete lack of recorded data) or pure-pre-animation (i.e., the impossibility to parameterized signs) by injecting elements of one approach into the other.

EMBR (Heloir and Kipp, 2009) was born as a tool for the synthetic animation of interactive virtual agents and was later employed in the generation of sign language animation (Heloir et al., 2011). It has been recently extended to support the playback of pre-animated facial movement (Kacorri and Huenerfauth, 2014), thus has become a mix between the two techniques. Its animation description language is not overspecialized for sign language, hence it offers more flexibility in the configuration of body postures. However, the SL generation is basically still performed

[1] `http://vh.cmp.uea.ac.uk/index.php/JASigning` – 23 Feb 2018

through a concatenation of poses and as such suffers of the same limitations as JASigning.
The player developed for the project ATLAS (Lombardo et al., 2010; Lombardo et al., 2011) is based on a repository of pre-animated clips. It supports sign parameterization in two ways. First, it allows for the overlapping of different animation tracks on top of the sign animation clip, allowing, for example, to control the animation of eyebrows and facial expressions independently from the arms. Second, it performs on-the-fly editing of the stored animation curves. For example, a sign can be relocated by applying an offset to the position of the hands for every frame of the clip. If the edits are applied on a limited range of deformation, the resulting motion looks still very natural and pleasant. However, the system lacks of any pure procedural synthesis of movement which is not exported from an animation editor.
The Paula system (Wolfe et al., 2011; McDonald et al., 2016), too, uses a multi-layered animation approach. The various layers can control the full body, with layers at higher priority overriding the animation of more basic layers. Each layer can playback pre-animated data as well as apply procedural control to body parts with routines tailored to support sign language animation (e.g., eyebrows adjustment and spine/torso rotation). Again, the range of possibilities of the player are limited to its database of motion clips and its hard-coded animation controllers, but can not be scripted via a high-level language.
These three above mentioned applications present a hybrid approach between the synthesis and the pre-animation. However, rather than deliberate architectural planning, it appears that one approach has been integrated into the other as a later attempt to overcome the limitations of an initial design choice. Additionally, none of those systems gives the possibility to independently drive different communicative channels on different time boundaries.
The only exeption we are aware of, is the design proposed by Filhol et al. (Filhol et al., 2017), where they explicitly designed from the very beginning a generative strategy accommodating both approaches together. A system that: i) is based on pre-animated signs, but ii) is ready to fall back on pure synthesis when pre-recording is not possible. Additionally, the system iii) do not enforce shared time boundaries on all the communicative channels.

3. The missing bottom-up system

Generic Sign synthesis platforms are designed to combine low-level (roughly, phonetic) features into larger pieces (lexical signs for the most part), stitched together in sequence to build utterances. For example, JASigning takes a string of units for input, labelled with glosses, each described with HamNoSys. In such phonetically inspired descriptions, a "Z" movement drawn in the vertical plane is composed with two horizontal strokes separated by a downward diagonal stroke. These individual strokes are part of the relatively small number of primitives of the language, but highly reusable for all sorts of movement descriptions. In this paper, we call such approach *bottom-up* animation, because it builds from the smallest possible features. Provided enough of these primitives—though by design, not necessarily plenty—the advantage is that one can describe everything by combination of the primary features. The work we propose here will also fall in this category.
On the other hand, the problem with bottom-up systems is that they inevitably render robotic animations. It comes from the fact that while humans may think of and describe certain movements as circles, straight movements or fixed orientations, actual human motion never follows its idealised geometric description. Animated as such, they do not look human, and there is no known generic way of taking the intended geometry and distorting it to look natural. The solution to provide naturalness is rather to make use of larger dedicated procedures or even full play-back of pre-recorded chunks, already implementing the human deviation from the idealised forms. This mostly advocates against bottom-up approaches all together, and for use of higher-level entries to avoid building complex movements from scratch.
Filhol et al. (Filhol et al., 2017) have recently reported testing such approach to naturalness for Sign Language with the Paula animation system. It works from:

- AZee, a language to specify linguistic input without fixed lexical signs and allowing more than merely listing Stokoe-style parameter values;
- the principle "the coarser the better", by which the larger the chunk of pre-animated data is, the better candidate it is for natural output;
- and the animation system Paula.

3.1. AZee input

AZee is a language to write parameterised signed forms for semantic functions. This can capture descriptions such as HamNoSys lexical entries (the fixed Stokoe-style description is the form; the gloss the meaning), but also more relational functions such as "not-but(X, Y)", meaning "not X but Y" and producing the form (say F) synchronising Y after X and a headshake and deep gaze in between, over a manual hold of X.
An AZee input for synthesis is typically a recursive nesting of semantic functions capturing the meaning of the production. For instance with the function above, plus "tree" and "wardrobe", one can build the following expression to mean "not a tree but a wardrobe":

not-but(tree(), wardrobe())

Evaluating this expression with an AZee parser produces a score, in this case F with blocks X and Y instantiated with the results of "tree" and "wardrobe" respectively, as illustrated in Fig. 1. The contents of a block is either:

- itself a score of the same recursive type, thus itself synchronising blocks on its nested time line (see outer boxes in the diagram);
- or a set of low-level constraints, which stops the recursion (grey-filled boxes).

The relevant constraints for this work are:

- placements of linguistically relevant body points (called *sites*) in target locations;
- bone orientations, e.g. orient normal vector of palm upright.

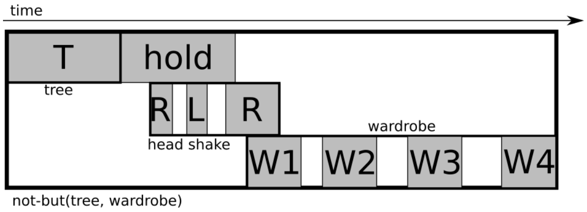


Figure 1: Example of AZee score.

For example, T in the diagram, representing the set of necessary and sufficient articulations constraints for the meaning "tree", is likely to include an orientation of the strong forearm up, a placement of the strong elbow on the weak palm, etc.

3.2. A top-down system

The idea Filhol & McDonald (Filhol et al., 2017) follow is to work with larger blocks rather than low-level features combined to produce synthesis. But it is not trivial to decide which blocks should be used. While the larger they are the more natural they look, the less feasible it becomes too because it is never possible to have everything prerecorded. Taking advantage of the recursive structure of AZee resulting scores, the authors address this problem in the following fashion.
What they do is start at the top level of the AZee score and work their way down the nested block structure until matches are found for blocks they have animations for (they call this a "short-cut"). At each level, if the full block is not matched, it is looked into and its constituents (sub-blocks) are layered on the animation score, each checked individually for a match, and so on. In contrast to what is done when building from small features to reach large blocks of utterances (bottom-up), this opposite approach can be called *top-down* animation.

3.3. What is missing

A top-down search for the most natural chunks guarantees that the highest usable blocks is used when appropriate, and never used when not. However, this system for the moment assumes a shortcut is possible at some point down the block structure, and that the bottom (low-level constraints specified in the non-breakable blocks) will not be met.
The problem is that actual SL generation systems involve blocks which cannot all be fully listed or recorded beforehand: infinite variation in continuous spaces, depiction, etc. AZee captures well these features with arbitrarily complex geometric expressions that are generally impossible to shortcut.
We propose to implement a bottom-up AZee animation system. No low-level AZee animation system exists yet, though it would ensure that anything can be animated from AZee, regardless of what we are ever able to shortcut. The goal is not to compete on the ground of naturalness since movements are bound to look robotic like all bottom-up systems. But its purpose could be to be used as the missing low-level fallback for a top-down system like Paula. This way we would take advantage of top-down shortcutting whenever possible, and still guarantee an output from AZee input when it is not, using bottom-up generation from low-level specifications.

4. Synthesising animations from AZee scores

One part of AZee which was missing so far–and described in this paper–is the implementation of the code realizing the skeletal poses for the keyframes delimiting the interpolation blocks.
The implementation of a high-quality gesture realizer encompasses a number of non trivial Computer Graphics and Animation techniques such as real-time rendering and shading, direct and inverse kinematics of complex kinematic chains, declaration and management of joint boundaries, collision detection, keyframe and timeline management, parameterizable interpolaton between animation curves, etc. All these features being available in the open source and liberally licensed Blender 3D editor[2], our AZee realizer is built in Python on top of the Blender API and could be viewed as an interface between AZee and Blender.

4.1. From AZee scores to an animation timeline

The AZee parser translates an AZee expression into a *score*, which is a set of timed intervals (blocks) whose boundaries are layed out on a timeline and whose contents is either:

- itself a nested score, such as "headshake" inside the "not-but" box in Fig. 1 (recursive case);
- a set of low-level constraints such as "R" inside the "headshake" box (base case).

The base case constraints include articulations (bone orientations and site placements) that can be thought as inverse kinematics (IK) problems in computer animation terms. Our goal is to translate those IK problems into (forward kinematics) joint rotations, and to position the full-body posture correctly on the final animation timeline.
This means create the right keyframes with the right set of constraints pulled and translated from the AZee resulting score, then relying on the system's native interpolation capability to fill the intermediate frames on the timeline. The first step to do so is to *flatten* the score from its nested and multi-linear structure so that it is projected on a single time line, as illustrated in Fig. 2.

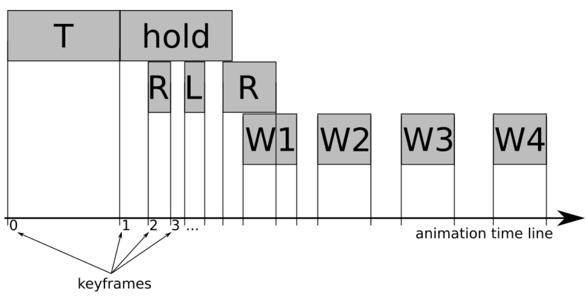


Figure 2: A *flattened* AZee score.

For every animation keyframe created on the timeline, we must copy all constraints that apply at that moment in time.

[2]https://www.blender.org – 23 Feb 2018

In our example, keyframes 0 and 1 will contain the set of constraints "T". Keyframes 2 and 3 will require not only a copy of "T" but also those from "R" since both apply at those dates on the flattened timeline.
At this point it becomes impossible to determine to provenance of the constraints packed in the keyframes. This is the step that would involve a crucial loss if we were aiming at naturalness because we will be relying on simple interpolation to fill all blanks. It is the reason for the expected robotic motion, and why it is not done by the top-down system presented above (working with the nested block structure). On the contrary, we are aiming at enabling synthesis from native constraints, so we accept this loss of naturalness to secure the possibility of an output.

4.2. Definitions: joints, sites, and IK-Problems

AZee defines the skeletal structure of a human signer in terms of *joints* and *sites*.
A **joint**, as in any 3D skeletal animation system, is a data structure characterized by:

- a *name*, unique for each joint of a skeleton;
- a *parent* joint, possibly *null* if the join is the root of the skeletal system;
- an *offset* from the parent joint. For the *root*, the offset represents the joint's absolute position relatively to the origin of the axes frame;
- a *rotation*, expressed by Euler angles for pose editing and then translated into a quaternion for better automatic interpolation;
- a set of *rotation limits*

  $$\{(R^x_{min}, R^x_{max}), (R^y_{min}, R^y_{max}), (R^z_{min}, R^z_{max})\}$$

  expressed as minimum and maximum rotation angle along the three Euler axes.

In this structure, a **bone** is essentially the segment connecting two joints with a parent-child relationship.

In addition to joints and bones, a **site** is defined as:

- a *name*, unique for each site;
- a *parent* joint from which the site depends;
- an *offset* from the parent joint.

Essentially, a site is a point in space whose absolute location depends from the global position and rotation of the parent joint. It is used to identify key points on the skin of the virtual human (e.g., the tip of a finger, or a corner of the mouth) that are used as reference for placement tasks. When a joint rotates, the linked bone, together with all the children joints and sites, change their absolute position.

The set of available bones and sites is described in two dedicated maps (*bones-map* and *sites-map*) thus separating the high-level namespace of the entities which can be addressed by AZee statements from the low-level hierarchical skeletal structure. This makes possible to seamlessly substitute the underlying character if a proper remapping is performed.

When AZee generates a keyframe on the timeline, the posture of the signer at that specific keyframe is described by a set of **IK-Problems**. An IK-Problem is defined by:

- an *IK-chain*, which is an ordered list of joints where the *start* of the chain, its first joint, is the parent of the second, the second is parent of the third, and so on. The last element of the chain can be a *site*, in which case it is called the *end-effector* of the chain;
- one optional *place constraint*. If present, it requires the end-effector of the chain (a site) to be positioned at a specific 3D point in space;
- a set of *rotation constraints*, possibly empty. Each rotation constraint requires either the *direction* or the *normal* of a bone to be *parallel* or *perpendicular* to a given 3D vector. The bone must be part of the IK-chain. Each bone can have at maximum two rotation constraints, in which case they will constrain independently both the direction and the normal of the bone.

AZee instantiates IK-Problems as an ordered list, taking into account their inter-dependencies. For instance with the set of constraints T in our previous example of the tree, at least two IK problems will be built, involving respectively:

- a placement of the weak hand at the target location of the tree;
- a placement of the strong elbow in contact with the weak hand palm.

The latter depending on the former, they will appear in this order in the list so that they can simply be applied in the given sequence. This way, the target placement of the strong elbow becomes a mere look-up of the current state of the avatar.
Determining the final pose of the virtual signer now requires solving, in the given order, all the IK-Problems defined on a given keyframe.

4.3. Solving single constraints

As introduced before, an IK-Problem is composed of one optional place-constraint and zero or more rotation-constraints.
The resolution of a place constraint yields to the resolution of the most classical of the IK problems, similar to robotics, where the end-effector of the ik-chain must be positioned in a 3D location. The IK solver calculates the rotation of all the joints of the chain.
Differently, solving a rotation-constraint for a bone means setting the absolute rotation in space of the bones's parent joint (e.g. to orient the forearm, we need to set the absolute rotation of the elbow). The IK-solver will determine the relative rotation of all the joints of the IK-chain up to the beginning of the chain. There are two cases, the first being when two rotation constraints are set at the same time on a bone. Forcing both the direction and the normal of a bone implies a unique possible absolute rotation for the bone's parent joint. In the second case, when only either the direction or the normal of a bone are set, the constraint is more relaxed and there are infinite solutions.
The three cases presented above (one for placing and two for rotation) can be individually solved by existing IK libraries. In our case, we use the iTaSC solver integrated in

```
1  def apply_ik_problem(skeleton, ik_prob):
2
3      # Outer iteration: apply all constraints of an IK-Problem.
4      iter_count = 0
5      while iter_count < MAX_ITER:
6          # Inner iterations: use the Blender IK to satisfy the single constraints.
7          if ik_prob.place_constr != null:
8              ik_prob.place_cstr.apply_to(skeleton)
9          for rot_constr in ik_prob.rot_cstr_list:
10             rot_cstr.apply_to(skeleton)
11
12         # Measure the ''distance'' between the current skeleton configuration
13         # and the the desired position/orientations
14         place_offset = ik_prob.place_cstr.offset_to_goal(skeleton) if ik_prob.place_cstr
               != null else 0.0
15
16         rot_offsets = [rot_cstr.offset_to_goal(skeleton) for rot_cstr in ik_prob.
               rot_cstr_list] if len(ik_prob.rot_cstr_list) > 0 else 0.0
17
18         # If all the distances are below the threshold, break the iteration
19         if place_offset < PLACE_THRSHLD and max(rot_offsets) < ROT_THRSH:
20             break
21
22         iter_count += 1
```

Listing 1: The algorithm describing the strategy to solve an IK-Problem.

the Blender software in SDLS mode. The iTaSC IK solver was originally developed by De Shutter et al. (De Schutter et al., 2007) and its integration in Blender is documented online[3].

4.4. Solving an IK-Problem

While solving the single constraints composing an IK-Problem is doable with off-the-shelf libraries, fulfilling the whole set of constraints in a single pose is not supported by the Blender iTaSC solver. In order to solve a whole IK-Problem into a final posture, we elaborated an algorithm whose pseudo code is shown in Listing 1.

The general strategy is to start by sequentially applying both the placement and all of the rotation constraints on the target skeleton. Each time a single constraint is applied, it is likely to break the position achieved by a previous constraint. Hence, we re-iterate the application of the single constraints until the final desired position is achieved.

We call this approach *two-level iteration*. The first *outer* level of the iteration begins at line 5. The second *inner* levels are immersed in the two invocation of the `apply_to` function at lines 8 and 10, where the basic (place or rotation) constraints composing an IK-Problem are solved using the Blender iTaSC solver. The `apply_to` function uses the Blender API to: i) create a *bone IK Constraint* on the Blender skeleton; ii) create a *target* object to drive the end-effector position or rotation; iii) trigger the execution of the iTaSC IK-solver, which solves the problem through a number of iterations (whose maximum value is set in the Blender properties); and iv) remove both the target object and the bone IK constraint.

[3] `https://wiki.blender.org/index.php/Dev:Source/GameEngine/RobotIKSolver` – 23 Feb 2018

The resolution takes automatically into account the joint rotation limits, which are applied as IK rotation limits in each Blender bone properties during a setup stage.

Line 14 computes the distance between the desired and the actual position of the site/end-effector.

Line 16 computes, for each joint in the IK-chain, the rotational distance between the current and the desired rotation. The rotational distance, which is an angle, is computed by first computing the quaternion needed to shift from the current to the desired rotation. The quaternion is then decomposed into an axis-angle representation and the angle in degrees returned.

Line 19: if all the distances are below the respective thresholds, the current pose is considered to be close enough to the desired one and the outer iteration breaks.

4.5. Implementation

The AZee animation system presented in this paper is implemented as add-on for Blender (v2.79). Figure 3 shows the GUI for AZee authors. On the right side, the author can move a virtual camera and see a 3D preview of the generated animation. A side panel shows buttons to setup the system, tune IK convergence parameters, and other debug flags. On the left, the author can insert AZee statements and execute them by clicking a button. At the bottom, a timeline marks where the AZee interpreter creates keyframes.

The current implementation operates on a prototype skeleton specifically developed for the AZee development. Future versions will address the problem of directly animating any imported human skeleton.

In our tests, we were able to solve an IK-Problem in a fraction of a second on an Intel(R) Core(TM) i7-3635QM CPU@2.40GHz (computation is limited to one core by the Blender architecture) and DDR3 RAM@1600Mhz

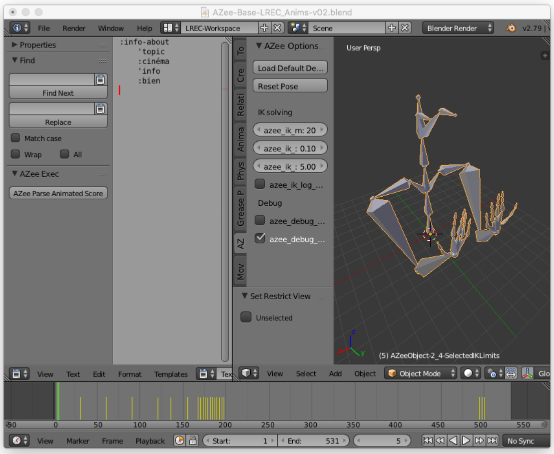


Figure 3: The AZee workspace in Blender.

Figure 4: Sign *tree* applied to a skinned skeleton.

with the following constant values: MAX_ITER=35, PLACE_THRSHLD=1cm, ROT_THRSHLD=5degs. As reported in detail in the next section, the computation time raises to several seconds when realizing a full signs or sentences. We are performing further tests in order to determine a trade-off between high precision (lower thresholds) and low realization time (lower number of iterations allowed).

5. Examples

In this section we present two working examples of the AZee animator. The first example is a static pose and aims at illustrating the mechanism of the IK resolution algorithm. The second example presents a more complex animated sentence and aims at illustrating the capability of AZee at interleaving communication channels.

5.1. Example1: tree

We report the results for the realization of the sign *arbre* (tree) in French Sign Language (see Figure 4). As shown in Figure 5-top, the sign requires the computation of 11 IK-Problems. The first two contain only a placement constraint: IK-Problem #0 asks for the palm of the left hand to be located in front of the signer, while IK-Problem #1 asks for the tip of the right elbow to touch the palm of the left hand. IK-Problem #2, the most complex, aligns at the same time the right forearm, hand, and thumb. The remaining IK-Problems straighten the four remaining fingers along the direction of the hand.

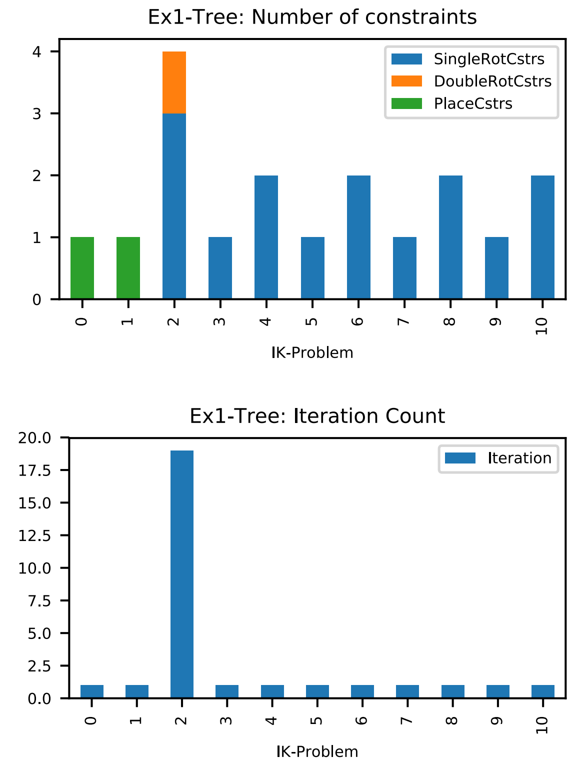


Figure 5: Sign *tree*. For each IK-Problem: (top) the count and the type of constraints, and (bottom) the number of iterations needed to solve the problem.

Figure 5-bottom reports the number of outer iterations needed to solve each IK-Problem. Ten out of the 11 IK-Problems solve with only 1 iteration, while IK-Problem #2 needs 19 iterations. This is expected, because problem #2 has 4 constraints and the application of each constraint is likely disrupting the orientation of the previous ones. In detail, IK-Problem #2 operates on an IK-chain going from the tip of the thumb to the elbow, and its constraints involve:

1. The orientation of the thumb tip, which must point outwards;
2. The direction of the normal of the thumb middle phalanx, which must face up;
3. The orientation of the hand, which must be aligned with the forearm;
4. The orientation of the forearm, which must point vertically up.

Figure 6 shows, for the sign *tree*, the side view of the position of the skeleton after a progressive number of iterations. Figure 7-top shows how the maximum rotational distance decreases as more iterations are executed. Figure 7-bottom shows the execution time for each iteration: for IK-Problem

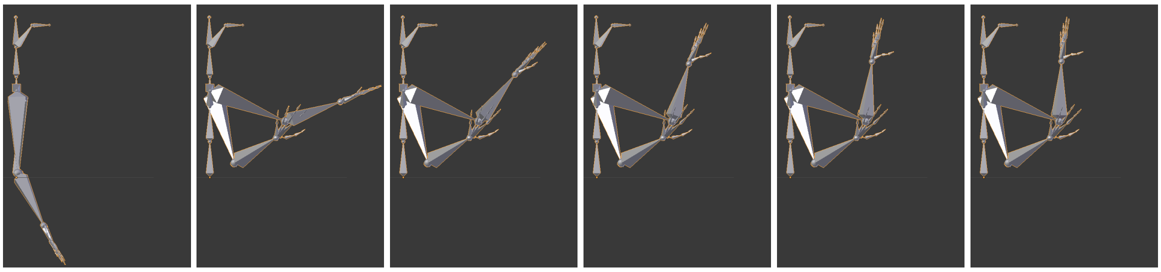


Figure 6: Sign *tree*: convergence to the final position from the resting pose (left) and after 1, 5, 10, 15, and 19 iterations.

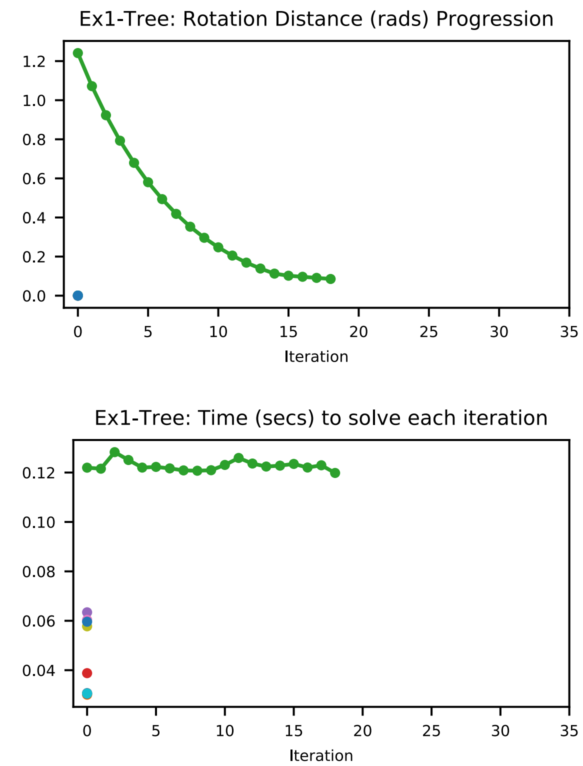


Figure 7: Test plot sizes.

#2 the execution time is stable at ca. 125ms for each iteration, while for all other IK-Problems the time lies between 25ms and 75ms for their first (and only) iteration. Overall, the interpretation of the sign pose took 2.76s to execute 29 iterations, with an average of 95ms per iterations (SD=39).

5.2. Example 2: cinema + good

The second example translates the sentence "The cinema is/was good":

```
:info-about
        'topic
        :cinema
        'info
        :bien
```

The top-level rule "info-about" produces a simple sequence of the two contained items, with a timed transition and an eye blink towards the end of the second item. IK is not be invoked for the final blink, and both contained items involve a classic sequence of manual postures at timed keyframes. At each keyframe, a number of IK problems is ordered.

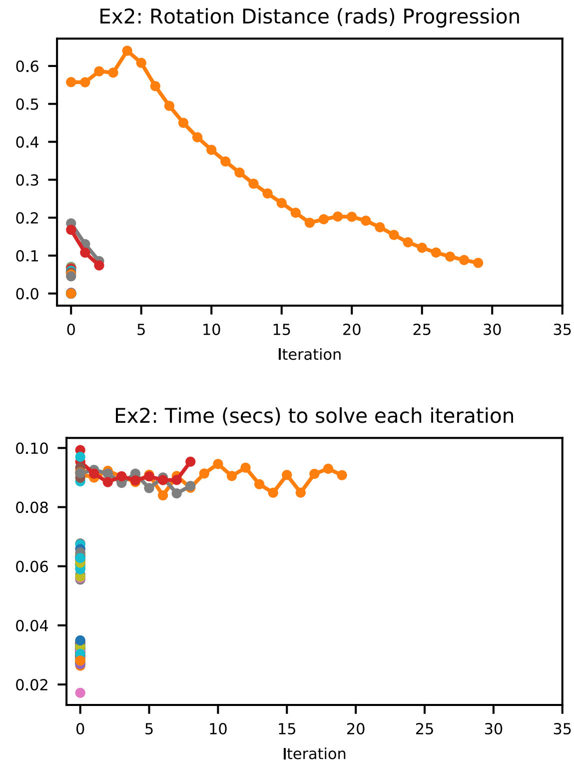


Figure 8: Test plot sizes.

The sentence required the resolution of 152 IK-Problems, of which 3 required multiple iterations. Figure 8-top shows the rotational distance progression, while Figure 8-bottom shows the execution time for each iteration. The hardest IK-Problem converged after 35 iterations. In general, all iterations are executed within ca. 100ms. Overall, the interpretation of the sentence took 9.97s to execute 185 iterations, with an average of 52ms per iteration (SD=26).

6. Conclusion

The work presented in this paper is the latest significant progress achieved in the implementation of a complete AZee realizer. This realizer is a step forward towards the implementation of a new generation of Sing Language synthesizers, allowing for the animation of different communicative channels which interleave on the timeline without the limitation of the hard boundaries dictated by lemma-based transcriptions.

Thanks to its integration in the Blender software, the realizer will allow for a streamlined and integrated generation of high quality rendered animations. Concerning the performances, from our tests it appears possible, on modern hardware, to generate the animation curves of a full sentence with limited delay and export them for playback on a turn-based interactive system.

The current implementation, still under test and development, behaves well in the realization of static poses, but still presents some glitches during the animation process, mainly occurring when the joints orientation get close to the boundaries of the IK rotation limits.

As described before, AZee separates the low-level description of the skeletal structure from an high-level namespace of *bones* and *sites*. Hence, AZee signs can be applied to different characters of any size and proportion without the need of facing retargeting problems. Given a new virtual interpreter, only the bones and sites name mapping must be updated. Certainly, the application to arbitrary avatars will suffer of self-compenetration issues, which must be addressed, for example, through the use of collision-prevention systems and the simulation of soft bodies.

This addresses one of the most frequent limitations of SL projects, which is the impossibility to interchange data between different virtual interpreters and database of animations. It makes AZee a good candidate language to create a database of animated signs which can be reused by any research group on SL, to animate their own avatar.
From the point of view of signal processing, AZee can be considered as a *lossy animation compression* format. Although focusing on the description and compression of sign language animations, it is possible to imagine Azee (or a variant of it) applied in the description of casual gesturing in general-purpose conversational agents. In the future, it might be possible to work on a system which automatically derives AZee descriptions from existing animation curves, allowing for the seamless transfer of sign repositories across different virtual interpreters.

7. Bibliographical References

De Schutter, J., De Laet, T., Rutgeerts, J., Decr, W., Smits, R., Aertbelin, E., Claes, K., and Bruyninckx, H. (2007). Constraint-based Task Specification and Estimation for Sensor-Based Robot Systems in the Presence of Geometric Uncertainty. *The International Journal of Robotics Research*, 26(5):433–455, May.

Efthimiou, E., Fontinea, S., Hanke, T., Glauert, J., Bowden, R., Braffort, A., Collet, C., Maragos, P., and Goudenove, F. (2010). Dicta-sign–sign language recognition, generation and modelling: a research effort with applications in deaf communication. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 80–83.

Filhol, M., McDonald, J., and Wolfe, R., (2017). *Synthesizing Sign Language by Connecting Linguistically Structured Descriptions to a Multi-track Animation System*, pages 27–40. Springer International Publishing, Cham.

Hanke, T. (2004). HamNoSys-representing sign language data in language resources and language processing contexts. In *LREC*, volume 4.

Heloir, A. and Kipp, M. (2009). EMBR - A Realtime Animation Engine for Interactive Embodied Agents. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents (IVA-09).*

Heloir, A., Nguyen, Q., and Kipp, M. (2011). Signing avatars: A feasibility study. In *The Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT), Dundee, Scotland, United Kingdom*.

Jennings, V., Elliott, R., Kennaway, R., and Glauert, J. (2010). Requirements for a signing avatar. In *Proceedings of the 4th LREC Workshop on the Representation and Processing of Sign Languages*, pages 133–136.

Kacorri, H. and Huenerfauth, M. (2014). Implementation and Evaluation of Animation Controls Sufficient for Conveying ASL Facial Expressions. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '14, pages 261–262, New York, NY, USA. ACM.

Lombardo, V., Nunnari, F., and Damiano, R. (2010). A virtual interpreter for the Italian sign language. In *Proceedings of the 10th international conference on Intelligent virtual agents*, IVA'10, pages 201–207, Berlin, Heidelberg. Springer-Verlag.

Lombardo, V., Battaglino, C., Damiano, R., and Nunnari, F. (2011). An Avatar-based Interface for the Italian Sign Language. In *Proceedings of the 2011 International Conference on Complex, Intelligent, and Software Intensive Systems*, CISIS '11, pages 589–594, Washington, DC, USA, June. IEEE Computer Society.

McDonald, J., Wolfe, R., Schnepp, J., Hochgesang, J., Jamrozik, D. G., Stumbo, M., Berke, L., Bialek, M., and Thomas, F. (2016). An automated technique for real-time production of lifelike animations of american sign language. *Universal Access in the Information Society*, 15(4):551–566, Nov.

Prillwitz, S. L., Leven, R., Zienert, H., Zienert, R., T.Hanke, and Henning, J. (1989). *HamNoSys Version 2.0.* International Studies on Sign Language and Communication of the Deaf.

Wolfe, R., McDonald, J., and Schnepp, J. C. (2011). Avatar to depict sign language: Building from reusable hand animation. January.

The Cologne Corpus of German Sign Language as L2 (C/CSL2): Current Development Stand

Alejandro Oviedo, Thomas Kaul, Leonid Klinner, Reiner Griebel
Department of Deaf Education, University of Cologne
{alejandro.oviedo, thomas.kaul, l.klinner, reiner.griebel}@uni-koeln.de

Abstract

The primary data of our learner corpus consists of 60 hours of videos produced by 350 L2/M2 learners of German Sign Language. The videos are monologues and dialogues that serve as tests at the end of each CEFR level (A1 to C1). An important part of the data includes videos produced by the same students at different times of their acquisition of the DGS over more than two years. Up to now, approximately 3% of the primary data have been transcribed (5,021 tokens; 281 lemmas). The corpus has already offered data for two studies on fluency in the DGS as L2. The biggest challenge we face is that most students refuse to grant permission for free access to their data in the corpus. Looking for solutions, we have made good experiences in obtaining those permits from students that have been directly linked to the research work related to the corpus.

Keywords: learner corpus, German Sign Language, fluency, resource availability.

1. Introduction

The University of Cologne offers courses of German Sign Language (DGS, henceforth) for about 900 L2 bachelor and master students at the Faculty of Human Sciences. Every year around 64 courses are taught from the beginners to the advanced CEFR-levels (Council of Europe, 2001) A1-C1. Upon completion of a level, students perform a series of tests of reception, interaction and production in DGS. Since summer 2015 a part of these tests has been video-recorded and archived with the aim of creating a learner corpus (Granger et al., 2015) of DGS as L2/M2[1] (Universität zu Köln, 2018).

2. Primary Data, Metadata

Currently, we have gathered about 60 hours of video, in more than 1,250 individual files (429 GB). They constitute the primary data of our learner corpus. Among the video files, about two thirds correspond to monologues (average duration 2.5 minutes) induced by an instruction (like "tell me what you did this week"), an illustration or a video. The rest of the files contains dialogues between the informant and a Deaf teacher (tests corresponding to the levels A1 to B2) or between two students (for level C1). Interaction videos have an average duration of 8 minutes. All test videos are archived, both approved and unapproved (students who do not pass an exam receive up to two new opportunities to present it. Every time, the test is video-recorded and archived). The data include a total of 350 informants (312 female and 38 male).

Metadata linked to these videos include age, gender and hearing status of the informants as well as the proof level and semester of data collection.

Part of the data is of a longitudinal nature, dealing with students who have visited our courses and presented the corresponding tests at various CEFR- levels between mid-2015 and the end of 2017. Currently, the most common settings correspond to A1-A2 and B1-B2. A small group of students have recorded videos from B1 to C1.

3. Transcription

So far, only a sample of 23 videos (17 monologues and 6 dialogues) has been transcribed and translated in ELAN (Osborn & Slotjes, 2008). Annotations consist of parent tiers containing German translations as well as ID-Glossing for every sign appearing in video[2]. The ID-Glossings are surrogates of the citation form of a sign and mostly assume the written form of the German word(s) historically related to the basic meaning of that sign. In the corpus, ID-Glossings are contrasted with the WebDGS (Universität zu Köln, 2008), a glossary of around 8,000 entries developed in our University. So far, we have registered 5,021 tokens and elaborated a list of 281 lemmas.

Special attention has been paid to segmentation. Following Hanke et al. (2012), transitions between two signs are not included as part of any of them. These moments of non-significant activity remain empty.

In addition to the annotation lines mentioned above, our transcription template includes five more series of tiers:

- **Deviations from the lexical norm:** These consist of variations observed in any manual parameter with respect to the native model (defined by the signs of our glossary WebDGS). There is one tier for each manual parameter (i.e. handshape, orientation,

[1]A small group of students has reported having a significant hearing loss. However, all of them have German as L1 and are therefore also included in the L2/M2 setting.

[2]A second line, ID-Glossing2, is available for cases in which each hand simultaneously articulates different signs.

location and movement). The activity of non-manual articulators can be transcribed using a tier for every articulator (e.g. head, eyebrows, nose, eyes, etc.). Most tiers mentioned above are attached to controlled vocabularies.

- **Elements of the utterance:** This annotation line include a controlled vocabulary comprising subject, object, nucleus of predication and predicate complements.
- **Fluencemes** (Götz, 2013), i.e., phenomena which interrupt the flow of lexical information. This group of lines includes pauses (empty and filled), repetitions and false starts.
- **Type of discourse:** It includes narration, explanation, description and argumentation (Grimes, 1975).
- **Paragraph limits**, in the sense of the border between the end of a thematic unit and the beginning of the next. (Longacre, 1979).

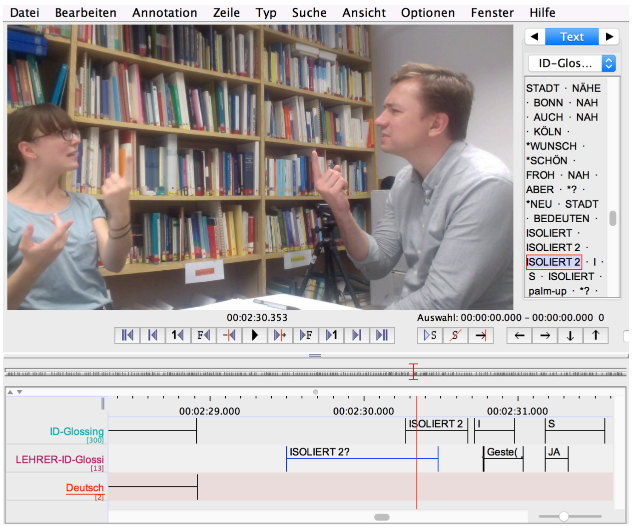


Figure 1. Screenshot of a C/CSL2 file in ELAN

4. Research Carried on C/CSL2 Data

4.1. The Problem of Assessing Fluency in DGS-Courses

Although little progress has been made in the elaboration of the corpus, available data have already served as the basis for an investigation on fluency, i.e., the capacity to produce complex chains of significant units with few interruptions (Fillmore, 1979). Fluency is seen as an inherent property of native discourse, and it offers a criterion for ranking learners at advanced L2 levels (Chambers, 1997). Since the introduction of the CEFR standards, the concept of fluency is a mandatory part of the curriculum. Regarding DGS, there is no research that allows us to know what makes a signed discourse fluent or not-fluent, what hampers the teaching and assessment of fluency in DGS courses.

4.2. First Survey

Looking for a solution to that problem, we have used our first corpus data (CEFR-level B1), as a basis for an experiment related with perceptions of fluency in L2 discourse (Kaul et al., 2017). The two selected videos shared the same stimulus and have a similar duration (3 min). But scores obtained differed: while the student of video 1 received the highest grade, the student of video 2 obtained the minimum one. The teachers who evaluated both tests, themselves part of our research team, thought that while video 1 showed a very fluent production, video 2 was less fluent. This difference had influenced the assigned grades.
Following a first analysis, the most relevant difference between both videos were the amount, form and distribution of pauses:

- Video 1 contained less pauses than video 2, and they were shorter (Video 1: Ø 743 ms vs. video 2: Ø 1326 ms).
- All pauses of video 1 were filled (with meaningless gestures, false starts or lengthened transitions between signs[3]), while most of the pauses in video 2 were empty (moments of no activity).
- Pauses in video 1 did not apparently have a fixed context of occurrence. Most pauses of video 2 appeared between sentences.

Additionally, both videos contained a similar number of repetitions and self-repairs. Regarding grammar, video 1 was free of errors, while video 2 contained some errors related to word order and non-manuals.
Based on these analysis, we developed a questionnaire in which we asked Deaf DGS users to rate both videos by assigning scores to seven items: fluency, pauses, repetitions, self-repairs, meaningless gestures, grammar correctness and intelligibility. Each aspect was rated in a 5-level-scale. 31 people answered the questionnaire after seeing each video twice.
Results: video 1 was rated as being more fluent than video 2. But both videos were rated as little fluent (between 3 and 4 at the given scale) and formal differences between both videos were perceived as quite smaller than expected. Correlations to fluency were found in both videos as highly significant for intelligibility and pauses and significant for repetitions. The regression analysis revealed that self-repairs (video 1) and intelligibility (video 2) were predictors of fluency. Summarizing, the indicators included in the questionnaire allowed to predict perceptions of fluency in video 1, but perceptions of fluency in video 2 seem to be partially determined by aspects not taken into consideration. For instance: Self-repairs played a role in video 1, but not in video 2. This could not be explained by our linguistic analysis.

[3]We did not have any reference to decide that the duration of a transition was normal or markedly longer than normal. Therefore, the decision about that was always taken in an impressionistic way by a Deaf researcher. Transitions marked as lengthened had an average duration of 900 ms.

4.3. Second Survey

The first study suggested that we were not considering the necessary indicators to determine why native users rate a non-native discourse as fluent or non-fluent. To advance in the solution of this problem, we carried on a study on possible indicators of fluency in native DGS discourse (Oviedo et al. to appear). We have recorded and transcribed eight videos produced by five Deaf native signers. Afterwards, we have asked a second group of users to indicate at what moments of the videos some type of interruption or disturbance in the signing flow was perceived. The indicated moments were marked and later used as orientation for the linguistic analysis.

Most marks corresponded to pauses, but the shape and distribution of them differed from the pauses we previously found. In native signing, the number of pauses was linked to the type of discourse. In argumentative or explicative fragments there were at least twice as many pauses as in narratives. Additionally, native signers gave empty pauses a regular distribution, namely in paragraph boundaries (Longacre, 1979). Filled pauses, however, did not have a predictable distribution. They seemed to serve as planning indicators. Finally, the pauses in L1 were, on average (Ø 550 ms), shorter than in our L2 sample.

Empty pauses of average duration were considered by the evaluators as fluent. On the contrary, almost all filled pauses were considered non-fluent. An exception was made when they marked paragraph boundaries and were accompanied by a clear interruption of the gaze contact (the gaze directed to a high point of the room).

Following the conclusions of the studies described above, we decided to add a new series of lines to our transcription template, in order to generate data regarding fluency in DGS as L2. These lines correspond to *fluencemes*, type of discourse and paragraph limits. When such data is available, we are planning to replicate the study by Kaul et al. (2017).

5. Challenges

5.1. No funding available

The work of organizing and transcribing the data has been done so far by the teachers and researchers of our work team in their free time. We do not yet have a specific funding to cover the costs of building up the corpus, so the project is progressing very slowly so far. We are working on applications for external funding.

5.2. Restricted access

The permissions to access and use the corpus remain, however, our biggest limitation. Before performing their tests, students are asked to sign a written form allowing the use of the videos as research data and / or as material for public presentations. So far, most students have granted authorization for restricted use of the videos (exclusively in our research group), but just a small group of 17 students have also authorized the full use of their data for both research and dissemination.

In order to expand access to the data, we are trying ways to convince our students to change their opinion about the necessary permits. Linking students actively in transcription and research work has proved to be promissory. We will summarize two experiences we have had in this sense:

Shortly after finishing the first study (Kaul et al., 2017), we asked a group of advanced DGS students to make some videos narrating in sign language the same story previously seen in a computer-animated film. We explained them that we wanted to analyze the videos with the collaboration of some Deaf teachers who did not belong to the university staff. The teachers, we added, wanted to improve the assessment of their own courses. All students agreed to permit this use of their videos. Subsequently, a group of external Deaf teachers held a workshop to seek criteria for evaluation and assessment of fluency in their DGS courses. Afterwards, our team discussed with the students the conclusions reached in the workshop and suggested strategies to increase their own signed fluency. All students expressed their understanding about providing data for studies that would improve the quality of teaching and authorized us to use their C/CSL2 data in an unrestricted manner.

Our second experience occurred within the framework of a linguistics seminar for M.A. students. We have presented them our corpus project and introduced them into the basics of sign language transcription. As final work of the course students could choose between a written essay or the transcription of a series of self-produced videos. The majority of students chose the second option. These transcriptions are currently in progress and should be incorporated into the corpus at the beginning of 2018. Most of this group of students have agreed that all their videos can be freely accessed. A repetition of this experience is planned for the next summer semester.

6. References

Chambers, F. (1997). What do we mean by fluency? In *System* 25(4), pp. 535–544.

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Press Syndicate of the University of Cambridge.

Crasborn, O., Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In Crasborn, O, Hanke, Th., Efthimiou, E., Zwitserlood, I., Thoutenhoofd, E. (Eds.). In *Sixth International Conference on Language Resources and Evaluation*, Paris: ELDA, pp. 39-43.

Fillmore, Ch. J. (1979). On fluency. In Fillmore, Ch.J., Kempler, D., Wang, W.Y. (Eds.). *Individual differences in language ability and language behavior*. New York: Academic Press, pp. 85–102.

Götz, S. (2013). *Fluency in Native and Nonnative English Speech*. Amsterdam: John Benjamins.

Granger, S., Guilquin, G., Meunier, F. (Eds.). (2015). *The Cambridge Handbook of Learner Corpus Research*. Cambridge University Press.

Grimes, J. E. (1975). *The thread of discourse*. The Hague: Mouton.

Hanke, T., Matthes, S., Regen, A., Worseck, S. (2012). Where does a sign start and end? Segmentation of

continuous signing, In *5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon Language Resources and Evaluation Conference (LREC)* (Istanbul), pp. 69–74.
Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. In *International Journal of Corpus Linguistics*, 15(1), pp. 106-131.
Kaul, Th., Oviedo, A., Griebel, R., Klinner, L., Prüfer, T., Krumpen, M. (2016). C/CSL2, The Cologne Corpus of Sign Language as a Second Language. Poster presented at TaLC 12, University of Giessen, 2/07/2016.
Kaul, Th., Oviedo, A., Griebel, R., Klinner, L., Glatz, A. (2017). How do Deaf native signers perceive fluency in adult L2/M2 learners of German Sign Language? Paper presented at the International Conference Fluency and Disfluency Across Languages and Language Varieties. Université Catholique de Louvain, 15/02/2017.
Longacre, R. E. (1979). The Paragraph as a Grammatical Unit. In Givón, T. (Ed.). *Syntax and Semantics*. New York: Academic Press, pp. 115–134.
Oviedo, A., Kaul, Th., Urbann, K., Griebel, R., Klinner, L. (to appear). Eine Annäherung zu den Pausen als Flüssigkeitsfaktoren in Deutscher Gebärdensprache als L1. In *Das Zeichen*, 32(108).

7. Language Resource References

Universität zu Köln (2018). C/CSL2. The Cologne Corpus of German Sign Language as L2/M2. Cologne: University of Cologne.
Universität zu Köln (2008). WebDGS: E-Learning für Gebärdensprache. Accessed at: http://webdgs.uni-koeln.de/ [20/02/2018]

The LESCO Corpus. Data for the Description of Costa Rican Sign Language

Alejandro Oviedo*, **Christian Ramírez Valerio****
*Department of Deaf Education, University of Cologne
**Doctoral Program in Social and Cultural Studies, University of Costa Rica
*alejandro.oviedo@uni-koeln.de, **xianlesco@yahoo.com

Abstract

The LESCO Corpus comprises transcriptions in Spanish glosses and translations into Spanish of videos of the Costa Rican Sign Language (known as LESCO). Transcriptions were made by a team of five Deaf LESCO users. The corpus was produced between 2011 and 2013 and has served as the basis for the development of a basic grammar and a dictionary of LESCO which are available online (www.cenarec-lesco.org. See Costa Rica, 2013a, b). The primary data includes 44 dialogues between two informants. Each film is composed of two or three video files. The corpus lasts approximately 2 hours and have a file volume of 12.6 GB. The sample includes 27 adults (average age: 32 years. 13 women and 14 men). Metadata correspond to age, sex, age of acquisition of LESCO, place of residence and hearing status. The material is protected under a Creative Commons BY-NC-SA license and its use may be requested to the Ministry of Education of Costa Rica (www.cenarec-lesco.org). The authors of this article were in charge of the entire project.

Keywords: sign language corpora, Costa Rican Sign Language, LESCO.

1. Introduction

In 2008, the government of Costa Rica decided to make a linguistic description of LESCO, the most widely used sign language in the country's urban areas, with the purpose of having a base for the subsequent creation of LESCO teaching materials for both L2 and L1 (for the schools for the Deaf).

The research team consisted of six people, five of them Deaf users of LESCO. The project lasted 27 months (between February 2011 and July 2013) and involved an investment of close to 250 thousand euros. A corpus, a grammar and a dictionary were the final products. All these products are freely distributed under a CC BY-NC-SA license. The grammar and the dictionary has been online since 2014 (Costa Rica, 2013a, b)

2. Sociolinguistics of the Costa Rican Deaf

2.1. Deaf Community

The single national association, ANASCOR (founded in the 1970s) brings together people from all over the country. The Deaf community is very active politically, and is present in many programs related to Deaf people, especially in the areas of education and interpreter training as well as in the legal field. LESCO was officially recognized in 2012.

There are only two schools for the Deaf in the country, which means that many of them attend regular schools. Schools for the Deaf, until a few years ago, excluded sign language from the classroom, but traditionally, students are allowed to sign outside of class hours. Since the beginning of the decade of 2010, there are some preschool bilingual programs (Spanish-LESCO) and more recently, a Spanish-LESCO bilingual classroom pilot program.

2.2. Old LESCO. New LESCO

In Costa Rican cities there are at least two different sign languages in use[1]: the old LESCO (apparently emerged throughout the first half of the twentieth century, with much influence from the Spanish Sign Language and used today by people over 60 years) and the new LESCO (emerged in the second half of the 1970s, with strong influence from the American Sign Language and used by people under 60 years of age (Woodward 1991).

Most people over 60, users of Old LESCO, are fluent users of New LESCO, unlike the younger people, for whom the Old LESCO is unintelligible. Given that the later objective of the project was the Deaf school population, the decision was made to limit the study to the New LESCO, so that informants older than 60 years were not included in the sample. In informal conversations between government officials and members of the Deaf community, it was also proposed to carry out a study to document and describe the Old LESCO. This study has not started yet, but some members of the Deaf community have been filming Deaf community elders[2].

Among the Deaf community of the country, the nominal phrases "Old LESCO" and "New LESCO" are used very rarely and exclusively in contexts where both languages are a topic of conversation. The sign in use that names the sign language in Costa Rica (see Figure 1) designates by antonomasia the language used by the new generations in the urban centers. That is, New LESCO. We follow that

[1] Following Woodward (1991) there would be two other sign languages in use among native American communities: The Bribri and the Brunca Sign Languages. No further study has corroborated the existence of these sign languages.

[2] Personal communication of Alejandro Oviedo with Costa Rican Deaf researchers Christian Ramírez Valerio and Alexander Hernández, in San José, May 2012.

use here and we will understand by LESCO the language that Woodward (1991) designated as New LESCO.

Figure 1: The sign LESCO

2.3. How many Users of (New) LESCO Are There?

To estimate the number of LESCO Deaf users, we resorted to a study similar to this one, carried out previously in Venezuela (Oviedo, 2004), whose social conditions were then comparable to those of Costa Rica. Oviedo (2004) estimated the local Deaf population crossing multiple data, such as the numbers of people with hearing disabilities, the number of members of Deaf associations, the school population, and the international percentage of babies born deaf. The estimate was 15,000 Deaf users of the Venezuelan Sign Language, 0.05% of the country's population at that time (26 MM). The present study includes Deaf users of LESCO living in the Greater Metropolitan Area (henceforth, GAM), where it is estimated that more than 60% of the population of Costa Rica lives (about 2.6 MM people (Costa Rica, 2011)). Based on the aforementioned percentage, we assumed that some 1,300 Deaf users of LESCO lived in the GAM. This would be the population of the present study.

3. Data Collection

3.1. Preparation and Application of the Surveys

To select the sample, 141[3] people between 18 and 60 years old were interviewed between the months of February and March 2011. Interviews were conducted using a written survey designed by a Deaf researcher. Questions were formulated and answered in LESCO, translated into Spanish and recorded in the interview form by the interviewer.
The people surveyed were selected in such a way that each of the main cities of the GAM was represented. The surveys were digitized and archived for further research.

3.2. Preparation for Video-Recording

The 141 surveys were subjected to a selection process, carried out by our Deaf researchers. Each potential informant was assigned a score of 1-10 (1 = minimum, 10 = maximum), according to four criteria: age of acquisition, attendance at a school for the deaf, frequency of contacts with other Deaf people and fluency in LESCO. This process allowed the pre-selection of 102 people.
These persons received invitations to come to the Ministry of Education for the filming sessions. They agreed to sign a document in which the Ministry of Education was allowed to use and eventually publish the data. The use of the videos was defined according to a Creative Commons BY-NC-SA license.

3.3. Video-Recording, Selection of the Films

3.3.1. Elicitation

Between March and May 2011, 196 filming sessions were made (2000 minutes, 34 hours, 482 video files, format 4:3). Texts were elicited and filmed according to the following scheme:

- An elicited narration (from a cartoon short film).
- A free narration (personal anecdotes).
- Unstructured dialogue between the informant and a Deaf researcher.
- Unstructured dialogue between two informants.
- Structured interview (based on a questionnaire)

All the filming took place in a dialogical situation, mostly with two cameras (Figure 2a), each of which took individual front shots of the participants. Some filming included a third camera directed at the informant's face (the second person present was always, in these cases, a Deaf researcher), with the intention of capturing facial activity in more detail (Figure 2b):

Figure 2a: A filming with two cameras

Figure 2b: A filming with three cameras

3.3.2. Video-Recording and Selection

Videos obtained were subjected to a selection process in which points were assigned (again 1 to 10) according to the following criteria:

- Fluency and intelligibility of discourse,

[3] I.e. 13% of the estimated local Deaf population.

- Technical quality of the video,
- Absence of information that could compromise third parties, and
- Closeness with the LESCO used by our Deaf researchers (since some informants produced a discourse that in the opinion of the researchers was not LESCO, but another signed system like signed Spanish or ASL).

Films receiving more than 7 points were selected. The final result was 44 films, in which a total of 27 informants appear (some of them appear in two, three or four videos). Figure 3 illustrates the distribution of these 27 informants - regarding age of acquisition and year of birth- among the interviewed 141 people:

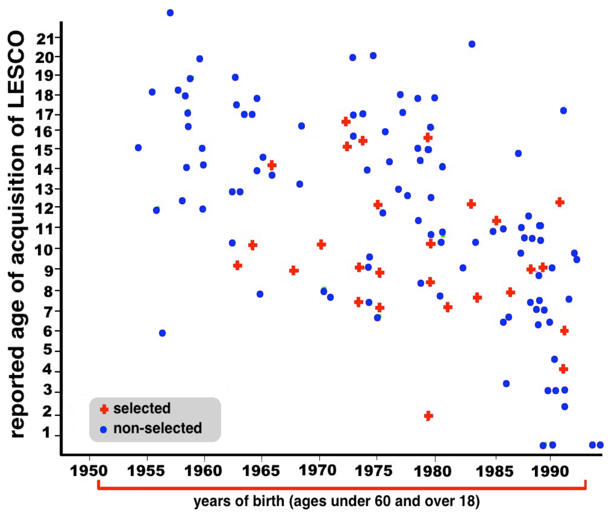


Figure 3. Data of informants

The red crosses in Figure 3 represent the people who were selected as informants for the LESCO Corpus. The blue dots represent the other 114 people who were interviewed in the selection process, but who were not selected as study informants. As can be seen, the relationship between age and age of LESCO acquisition of the selected people corresponds to the sample trend. The younger the age, the younger the age of acquisition. In the sample (141 people interviewed) the average age of acquisition was 11.27 years. Among the 27 selected it is 10.36 years.
Determining the age of acquisition was a complex procedure, since it depended on the assertions of the informant himself, based on childhood memories. In each case, the school and family history that each person reported was also taken into account. Among the 27 informants selected there were several who report relatively late ages of acquisition (between 12 and 17 years), but at the same time they report having Deaf relatives (siblings, uncles, cousins). In such cases, it is legitimate to assume that the actual acquisition ages would be lower than those reported. This would eventually allow us to further reduce the acquisition age of the selected group.
Finally, it should be noted that four Deaf CODAs were included in the initial sample. They correspond to the four blue points near the right end of the horizontal line in Figure 3 above. These people are assigned an acquisition age of 0 years. However, we excluded these four people as informants, since two of them were underage and the other two produced discourse considered by our team as signed Spanish.

4. Transcription

4.1. Conventions, general description

Corpus files contain a Spanish translation and an ID-Glossing line for each signer that appears in the video. Given glosses generally assumed the written form of the Spanish word(s), which the Deaf community relates to the basic meaning of that sign. Transitions between signs were not marked, but were included as part of the end of a sign and/or the beginning of the next sign. Therefore, only when pauses occur, the annotation line is interrupted. When the pauses were long (400 ms or more), the word PAUSED (pause) was included to fill the space in the ID-Glossing line. Finally, there was a series of annotation lines for manual parameters (i.e. handshape, orientation, location and movement) as well as for the non-manual articulators (head, shoulder, eyebrows, nose, eyes, lips, etc.). Figure 4 illustrates the transcription of a corpus file in ELAN:

Figure 4. Screenshot of a corpus file

4.2. Lemmatization

Before the realization of our study, at least two LESCO vocabularies were published (Bravo, 1979; López Gracioso, 1992). They were lists of Spanish glosses, arranged alphabetically and illustrated by means of a drawing or a photo. These works were our first references in the process of lemmatization (Johnston, 2010). Many signs in the corpus, however, did not appear in any of the works mentioned. In those cases, members of the research team assigned temporary glosses. Once the signs were transcribed in the corpus, we exported in ELAN the transcriptions as lists of words and compared the occurrences, taking into account basic form, changes of form observed in the corpus, meaning and use. In this process, errors were corrected or new lemmas were added. The result was a list of 1,541 lemmas and almost 14,000 tokens.

5. Some notes about the grammar and the dictionary of LESCO

5.1. The Grammar

The grammar of LESCO (Costa Rica, 2013b) consists of a text written in Spanish, divided into four major parts: phonetic-phonological, morphological, syntactic and discursive levels. In each of them, a series of subtopics offers explanations on more concrete aspects. For example, in the case of morphology, it provides information about the meanings and forms that the repetition of the lexical root can carry. In each case, the explanation is illustrated with numerous examples of the corpus, which can be observed in both photos and videos.

5.2. The dictionary

The dictionary of LESCO (Costa Rica, 2013a) was elaborated from the lemmas defined in the process of elaboration of the corpus. Out of the 1,541 lemmas defined, about 960 were selected for the dictionary. In some cases, certain semantic fields were not saturated with the signs found in the corpus. It was the case, for example, of the signs corresponding to colors. To complete them, we reviewed the videos that had not been selected for transcription. In this review, many of the signs sought were found and included in the dictionary. If after this process some semantic field still remained unsaturated (as with the signs corresponding to Costa Rican provinces and cities) we consulted members of the Deaf community of San José about the missing signs. Once a consensus was reached, a neutral form of the missing sign was filmed and a new lemma was defined. With this procedure it was possible to complete a list of 1,041 signs for the dictionary. Figure 5 illustrates an example of this (the entry for LESCO in the dictionary):

Figure 5. Screenshot of a dictionary´s entry

Each of those 1,041 lemmas has an entry in the dictionary. The signs can be found through three search criteria:

- The Spanish gloss (in alphabetical order),
- The manual configuration of the active hand (ordered according to the number of active fingers)
- Through a thematic index.

Each entry comprises a video of the neutral form ("forma neutra") as well as one or two videos taken from the corpus, in which examples of use of the sign appear. These examples are glossed and translated for ease of use. Additionally, the entry contains some information about the grammar and the meaning of the sign.

As far as we know, our work on LESCO was the second corpus-based description of a signed language in Hispanic Latin America. A previous experienced was carried out in Colombia between 2000 and 2005 (Oviedo, 2001; Colombia, 2006).

6. References

Bravo, E. (1979). *Hacia una nueva forma de comunicación con el sordo.* San José: MEP.

Colombia. Ministerio de Educación Nacional. Instituto Nacional para Sordos (INSOR) / Instituto Caro y Cuervo (ICyC) (2006). *Diccionario Básico de la Lengua de Señas Colombiana*. Bogotá, D.C.: Ministerio de Educación Nacional, Instituto Nacional para Sordos, INSOR.

Costa Rica. (2004). *La discapacidad en Costa Rica: situación actual y perspectivas.* San José: Organización Panamericana de Salud y Ministerio de Salud.

Costa Rica. Instituto Nacional de Estadística y Censos (2011). *Resultados Generales Censo 2011*. San José: INEC.

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. In *International Journal of Corpus Linguistics,* 15(1), pp. 106-131.

López Gracioso, D.M. (1992). *Comuniquémonos mejor. Diccionario Ilustrado de Lengua de Señas Costarricense*. San José: Deisa Internacional.

Oviedo, A. (2004). *A Study on Classifiers in Venezuelan Sign Language*. Seedorf: Signum, 2004.

Oviedo, A. (2001). *Apuntes para una gramática de la Lengua de Señas Colombiana*. Cali: Universidad del Valle/INSOR.

Woodward, J. (1991). Sign language varieties in Costa Rica. In *Sign Language Studies* (20), pp. 329-334.

7. Language Resource References

Costa Rica. CENAREC (2013a). Diccionario Básico de la LESCO. Accessed at: http://cenarec-lesco.org/DiccionarioLESCO.php [21/02/2018]

Costa Rica. CENAREC (2013b). Gramática Básica de la LESCO. Accessed at: http://www.cenarec-lesco.org/index.php/ [21/02/2018]

Crowdsourcing for the Swedish Sign Language Dictionary

Nikolaus Riemer Kankkonen, Thomas Björkstrand, Johanna Mesch, Carl Börstell
Department of Linguistics, Stockholm University
S-106 91 Stockholm, Sweden
{nikolaus.kankkonen, bjorkstrand, johanna.mesch, calle}@ling.su.se

Abstract

In this paper, we describe how we are actively using the Swedish Sign Language (SSL) community in collecting and documenting signs and lexical variation for our language resources, particularly the online Swedish Sign Language Dictionary (SSLD). Apart from using the SSL Corpus as a source of input for new signs and lexical variation in the SSLD, we also involve the community in two ways: first, we interact with SSL signers directly at various venues, collecting signs and judgments about signs; second, we discuss sign usage, lexical variation, and sign formation with SSL signers on social media, particularly through a Facebook group in which we both actively engage in and monitor discussions about SSL. Through these channels, we are able to get direct feedback on our language documentation work and improve on what has become the main lexicographic resource for SSL. We describe the process of simultaneously using corpus data, judgment and elicitation data, and crowdsourcing and discussion groups for enhancing the SSLD, and give examples of findings pertaining to lexical variation resulting from this work.

Keywords: Swedisch Sign Language, lexicography, variation, sociolinguistics

1. Introduction

The Swedish Sign Language Dictionary (SSLD) (Svenskt teckenspråkslexikon, 2018) has been the main lexical database for Swedish Sign Language (SSL) since 2008. Initially constructed as an online video representation of an earlier printed dictionary (Hedberg et al., 1998), it soon grew to a limited resource following criteria of the printed dictionary.

An important aspect of any language resource is that it is representative of the language or register it covers. In the case of resources also serving as language documentation, this is perhaps even more important. There are currently two main language resources of SSL publicly available, a dictionary and a corpus (Mesch et al., 2012; Mesch and Wallin, 2012), both of which also serve as a form of language documentation for SSL. The first is the Swedish Sign Language Dictionary (SSLD), which is an online video dictionary, currently containing 17,310 entries, and in some cases sentence examples (available for around 22% of the entries, in total 3,944 sentence examples). Each dictionary entry is represented by a video of the sign (or phrase), a Swedish translation, phonological information, and internal cross-links to phonologically or semantically equivalent signs – i.e. homophones and synonyms (Svenskt teckenspråkslexikon, 2018). The second resource is the Swedish Sign Language Corpus (SSLC), which consists of 24 hours of video data (conversations, narratives, and presentations) from 42 different signers collected in the years 2009–2011 (Mesch et al., 2012; Mesch, 2018).

Although both the SSLD and the SSLC are designed and maintained by the Sign Language Section at the Department of Linguistics, Stockholm University, they have initially been set up independently from each other. This has led to some discrepancies between the functions of the two resources. However, the two projects have started to converge in the last year, and now they are already supporting each other. For example, sign tokens in the SSLC are annotated with the SSLD ID number (where matching is possible), signs are added to the SSLD as they appear in the SSLC, and the SSLD online interface sorts synonym search hits according to SSLC frequencies calculated through the SSLC lexical frequency interface (Börstell and Östling, 2016). However, seeing as the SSLC currently contains only around 90,000 sign tokens, and there was no explicit lexical variation elicitation task during the collection of the corpus data (Stamp et al., 2014), many sign synonyms or form variations cannot yet be investigated solely with the use of corpus data.[1] A set of 90,000 sign tokens is fairly large for being a sign language corpus, but small when compared to spoken language corpora used to investigate variation systematically. Instead, we make use of crowdsourcing to overcome some of these obstacles. In this paper, we present our methods for involving the language community to collect new signs and to gather data on familiarity with and variation within sign synonyms – that is, crowdsourcing to improve our language resources, specifically the SSLD. Thus, we include the community in enhancing the functionality of the resources later used by the community itself.

2. What Is the Sign for X?

The SSLD is an important resource for many different groups within the SSL community. It is used by SSL students, interpreters, and also Deaf signers as the go-to reference for looking up signs and sign variants in their language. Two common questions that are asked are: a) *what is the sign for X?*; and b) *which of the sign variants for X should I use?* Regarding the first question, it is an issue for language documentation: we need to find, document, and distribute knowledge about the signs that are used by the community. As for the second question, there is sometimes a general "hunch" as to which variant from a set of signs with more or less similar form is used by which sociolectal or dialectal group (e.g., *"This is an older sign"*, or, *"This*

[1] Approximately half of the collected 24 hours of data have been annotated, thus far.

sign is used mostly in Stockholm"), but we aim to provide a research-based and data-driven dictionary, which means we need more data to support such claims. In some cases, we may use the SSLC data and its frequency tool (Börstell and Östling, 2018) to compare the relative frequencies of two competing sign variants. An example of this is shown in Figures 1–2 with the sign TIO ('ten') in two different forms, a one-handed (Figure 3) and an older two-handed form (Figure 4).[2]

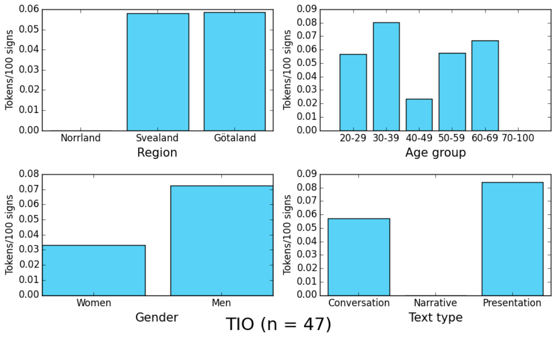


Figure 1: The relative frequencies of the one-handed sign TIO ('ten') – SSLD ID 4475

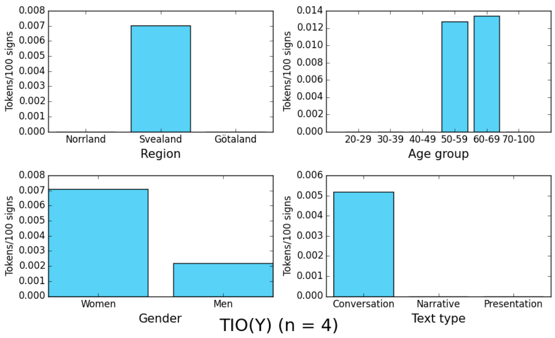


Figure 2: The relative frequencies of the two-handed sign TIO(Y) ('ten') – SSLD ID 11951

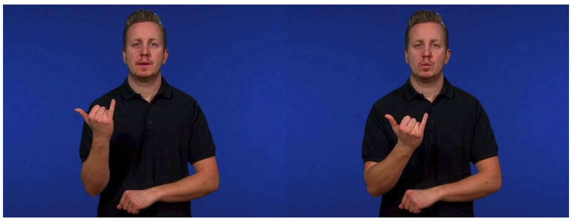


Figure 3: The one-handed sign TIO (SSLD ID 4475) ('ten')

As the figures show, the fact that TIO(Y) is an older sign than TIO is visible in their respective distributions: TIO

[2]The ID numbers in the figure captions are linked to the sign entry in the SSLD.

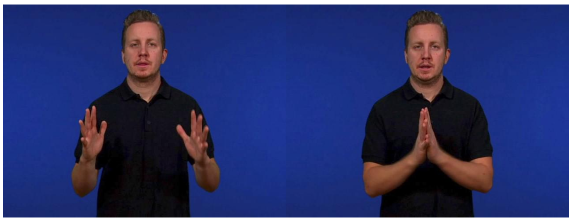


Figure 4: The two-handed sign TIO(Y) (SSLD ID 11951) ('ten')

is more common overall, and TIO(Y) is restricted to older signers. However, already here we are dealing with quite few data points (only four tokens for TIO(Y)), which is where our crowdsourcing comes into play.

We have two main crowdsourcing strategies: the first is a Facebook group entitled *Teckenspråkslexikon* ('sign language dictionary') administrated by our dictionary team; the second is our annual participation in *Dövas Dag* ('the Deaf Day'), which is a national convention for Deaf organizations and Deaf community businesses and activities.

3. The Facebook Group

Our Facebook group *Teckenspråkslexikon* was started in October 2014, after that year's Deaf Day, by the dictionary team. It currently has 2,642 members – see Figure 5. Figure 6 shows the monthly increase of members in the group from its start until now (October 2014–January 2018). It is visible here that the number of members greatly increases around two points during the year: September–October and January–February. We expect that this is due to two specific events. First, the Deaf Day is organized in September each year, when the SSLD team has informed about their work there. Second, our sign language courses at the Department of Linguistics, Stockholm University, start in September and January. The Facebook group has a diverse set of members, consisting of Deaf, hard-of-hearing, and hearing individuals, who are involved in the community in different ways (Deaf, Codas, interpreters, and friends and relatives of Deaf people, etc.).

The group is based around members asking for and discussing different signs, mainly under the premise that the meaning is not found in the online dictionary. Members then interact in different ways, discussing which signs are to be used, or the difference between sign variants. This interaction takes place both in written Swedish and in SSL by members uploading their own video comments, thanks to the video comment function in Facebook. In the last two years (i.e., January 2016 to January 2018), there have been 593 posts and a total of 5,817 interactions (original posts and comments combined) – see Figure 7. The peak phases of the period show that the posts and comments appear most in synchrony with peaks in member expansion – i.e., September–October and January–February (cf. Figure 6).

Table 1 shows the distribution of post types in the Facebook group from January 2016 to January 2018. As the table shows, 84% of the all posts consist of questions about

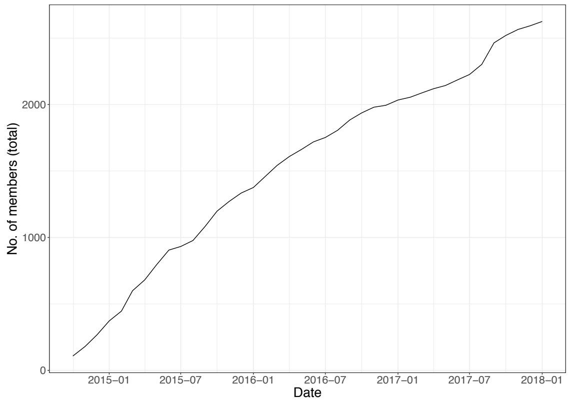


Figure 5: The number of total members in the Facebook group between October 2014 and January 2018

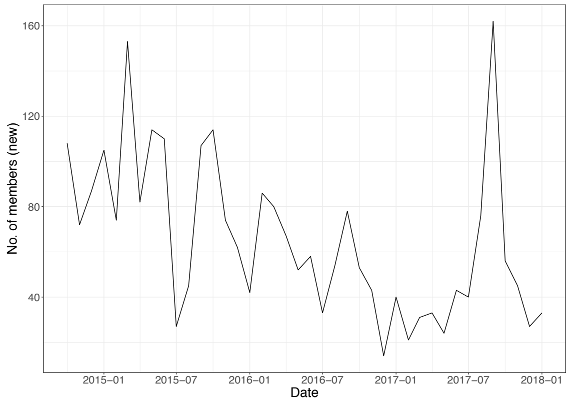


Figure 6: The number of new members joining the Facebook group per month between October 2014 and January 2018

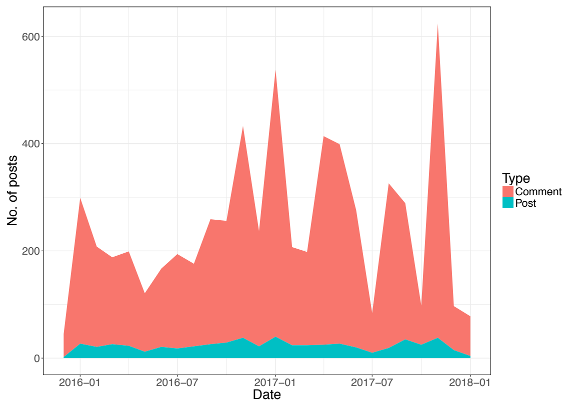


Figure 7: Amount of posts and comments of the Facebook group members between January 2016 and January 2018

signs. This includes questions about signs, such as *how do you sign X?* (53%), or name signs, such as *is there a sign for person Y?* (20%), but also general questions about the usage (10%) and etymology (<0.2%) of specific signs. 12% of the posts are about information, of which half are people giving information about something related to SSL or the SSLD (e.g., new features in the SSLD, or language resources), and the other half are requests for information about some topic (e.g., sign language courses or tools). Finally, 4% of the posts are not directly related to the SSLD or SSL, and are thus categorized as *Other*.

Post type	Number	%
Question	498	84%
– *sign*	*315*	*53%*
– *name sign*	*121*	*20%*
– *usage*	*61*	*10%*
– *etymology*	*1*	*<0.2%*
Information	72	12%
– *giving*	*36*	*6%*
– *searching*	*36*	*6%*
Other	23	4%
	593	100%

Table 1: Number of posts by type in the Facebook group (December 2015–January 2018

Of the posts concerning questions about signs, approximately half of the posts concern signs already found in the SSLD, whereas the other half do not. However, there are some signs requested that were already in pre-published stage of the SSLD, and many were added to the SSLD after the request was made, some of which were based on suggestions in the comments. Through the comment section, it is possible to follow the discussions of the community members and their reactions and ideas about signs in the SSLD or signs suggested in video comments. However, there are only few discussions about whether a sign entry in the SSLD is correct or not, and in those cases, approximately 90% of the commentators think that the SSLD sign entry is good.

Other discussions about signs may be about a meaning or the lexical category of a sign, for example TO-PROGRAM (verb) vs. PROGRAMMER (a person who writes computer software). Especially when signs for new concepts (e.g., technology) are demonstrated in video comments, we can, based on the reaction and informal ratings (e.g., reaction buttons), choose to include the sign in the dictionary, as part of language documentation and a way of enhancing the language resource for the public. Thus, the Facebook group helps us answer both questions (partially) – i.e., both *what* new signs there are, and *who* uses them. Many new sign entries have been added to the SSLD as a direct consequence of their being demonstrated in the Facebook group.

4. Direct Contact with the SSL Community

During our participation at the last Deaf Day in September 2017, we had devised a questionnaire in Google Forms with the intention of collecting variation data from community members. The questionnaire started with a set of background questions (i.e., signer metadata) followed by

25 questions of the type *"How do you sign X?"*, with all the documented sign variants available as animated `.gif` files.[3] Deaf attendees at the convention would participate in our study by responding to the questionnaire on a computer. For each item in the 25-item concept list, the signer would first be presented with the Swedish translation of the word and then demonstrate their preferred sign for the meaning specified before the questionnaire administrator would continue to a subsequent page showing the animated `.gifs`. At this stage, the signer could choose to mark several signs as possible variants they would use. We collected data from 26 signers (12 female, 14 male; mean age 47; median age 49).[4] Their responses were compiled and sorted by signer metadata into a pilot study for evaluating the usefulness of the questionnaire. Data compilation was done with the statistical language R (R Core Team, 2015) and the data were plotted with the `ggplot2` package (Wickham, 2009).
In Figures 8–12 below, we illustrate the responses from our 26 Deaf primary SSL users for the meanings TIO ('ten'), NITTIO ('ninety'), and TORSDAG ('Thursday') across age groups (bins show decade of birth).

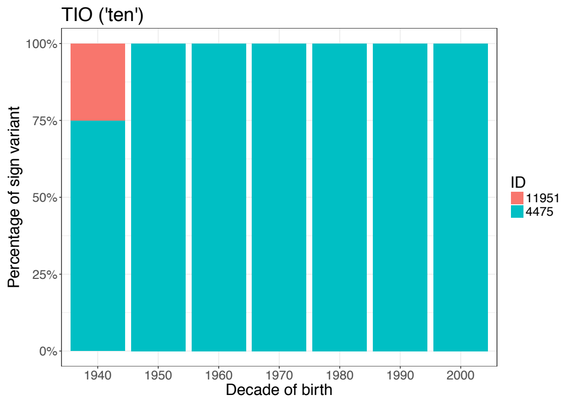


Figure 8: The relative distribution of the signs TIO (4475) and TIO(Y) (11951) ('ten')

Figure 8 confirms the pattern found in Figures 1–2, namely that TIO(Y) is less common overall, and that it is mainly confined to the usage of older signers (only present in the oldest age group here).
From Figure 9, we see that the sign form NITTIO(E) is becoming more popular over time, which we again would predict from knowing that NITTIO(4) is the older sign variant, and also a suppletive numeral by not being based on the sign NIO ('nine') as other tens.
Lastly, Figure 12 gives us the clearest example of a form change over time. The sign TORSDAG(L) is mostly used by signers born before 1980, and TORSDAG(Lb) mostly by

[3] We have converted all sign videos (`.mp4`) in the SSLD to `.gif` format with the intention to make these available in the online database for the public in a future release. The purpose of adding `.gif` files is to allow for sharing signs in, e.g., social media more easily. However, the original `.mp4` videos will remain the primary format in the sign videos in the SSLD.

[4] In total, 32 people responded to our questionnaire. Here, we only report the results of the respondents identifying as Deaf.

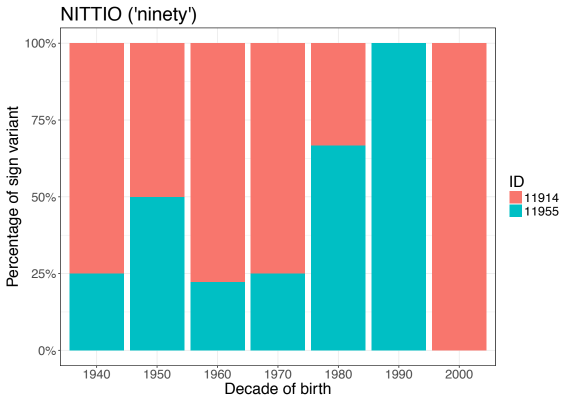


Figure 9: The relative distribution of the signs NITTIO(4) (11914) and NITTIO(E) (11955) ('ninety')

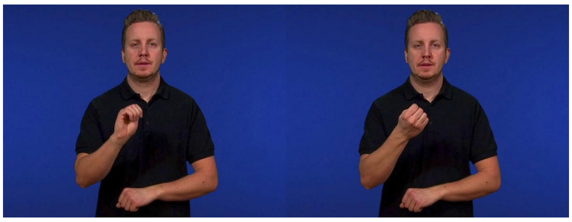


Figure 10: The sign NITTIO(E) (11955) ('ninety')

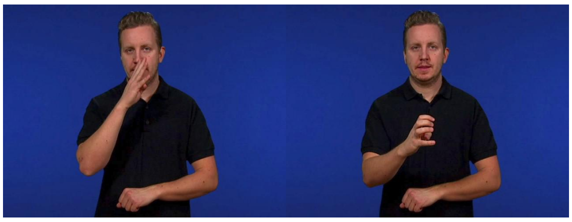


Figure 11: The sign NITTIO(4) (11914) ('ninety')

signers born after 1970, with the 1970s and 1980s as the transitional decades.
Here, we have shown three examples of meanings for which there was an suspected diachronic change in which sign variants are used, and the hypotheses about their distribution and change in distribution are tentatively confirmed by our data. Although these three examples give us an indication of patterns, we would need to gather much more data in order for it to be reliable and thus useful. We have the intention of continuing with the questionnaire type data collection, but this time entirely online in order to reach a larger set of community members more efficiently.

5. Conclusions

In this paper, we have described two of the main methods that we use in order to crowdsource data about SSL directly from the SSL community.
In the case of our Facebook group, we are able to collect

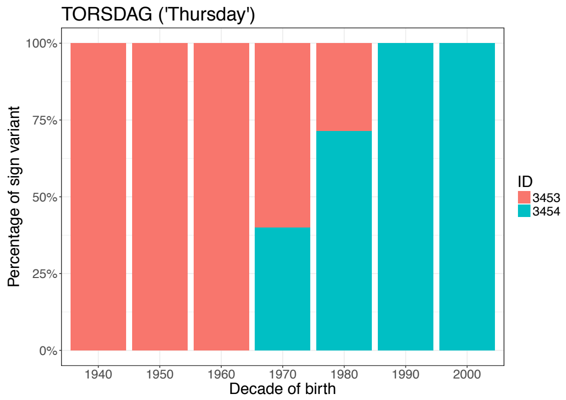


Figure 12: The relative distribution of the signs TORSDAG(L) (3453) and TORSDAG(Lb) (3454) ('Thursday')

Figure 13: The sign TORSDAG(L) (3453) ('Thursday')

Figure 14: The sign TORSDAG(Lb) (3454) ('Thursday')

signs directly, disseminate information about (new) signs to various subgroups in the community (Deaf, hard-of-hearing, hearing) – both ourselves and in interaction with other community members – and also to collect judgments about signs already in the SSLD or signs that could be included. The on-going documentation process enables us to quickly add new signs in consultation with the community members when needed, and also update or edit entries accordingly.

With our direct interaction with the community members at Deaf events (e.g., the Deaf Day), we are able to collect data with more extensive metadata about each signer, target specific individuals or groups that we need for our documentation work, and ask more qualitative questions about signs and sign usage. Thus, it complements the more passive (and massive) interaction taking place in our Facebook group. The results from our pilot study questionnaire shows that it could be a useful method for moving on to a larger online variation study, using a similar type of questionnaire with respondents being able to provide data from their own homes rather than requiring a direct interaction. This would further utilize the benefits of crowdsourcing, which can lead to a lot of new data in a very short time.

Crowdsourcing for the SSLD is a useful and rapid method for enhancing our language resource. It is especially convenient using the online community (e.g., our Facebook group) since it easily targets a large group of community members simultaneously. However, data collected through offline methods give may provide a more qualitative approach on sign variants, interacting with individuals directly, and allows for better control over signer metadata and responses, by being manually annotated – aside from giving the dictionary team an opportunity to personally meet and interact with the deaf community, which is an important aspect of any language documentation work.

6. Acknowledgments

We wish to thank the members of the SSL community who, through channels like our Facebook group and the Deaf Day, help us enrich and enhance our language resources part of our documentation and research of Swedish Sign Language. We are also grateful to the reviewers for helpful comments on a previous version of this paper.

7. Bibliographical References

Börstell, C. and Östling, R. (2016). Visualizing lects in a sign language corpus: Mining lexical variation data in lects of Swedish Sign Language. In Eleni Efthimiou, et al., editors, *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*, pages 13–18, Paris. European Language Resources Association (ELRA).

Hedberg, T., Lindström, M., Lyxell, T., Mumtaz, S. A., and Tersmeden-Lundquist, E. (1998). *Svenskt teckenspråkslexikon*. Sveriges dövas riksförbund (SDR), Leksand.

Mesch, J. and Wallin, L. (2012). From meaning to signs and back: Lexicography and the Swedish Sign Language Corpus. In Onno Crasborn, et al., editors, *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon [Language Resources and Evaluation Conference (LREC)]*, pages 123–126, Paris. European Language Resources Association (ELRA).

Mesch, J., Wallin, L., and Björkstrand, T. (2012). Sign Language Resources in Sweden: Dictionary and Corpus. In Onno Crasborn, et al., editors, *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon [Language Resources and Evaluation Conference (LREC)]*, pages 127–130, Paris. European Language Resources Association (ELRA).

R Core Team, (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Stamp, R., Schembri, A., Fenlon, J., Rentelis, R., Woll, B., and Cormier, K. (2014). Lexical variation and change in British sign language. *PLoS ONE*, 9(4).

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

8. Language Resource References

Börstell, C. and Östling, R. (2018). SSL-lects: An online interface for plotting lexical variation data from the Swedish Sign Language Corpus. Department of Linguistics, Stockholm University. mumin.ling.su.se/cgi-bin/ssllects.py.

Mesch, J., Wallin, L., Nilsson, A.-L., and Bergman, B. (2012). Dataset. Swedish Sign Language Corpus project 2009–2011 (version 1).

Mesch, J. (2018). Annotated files for the Swedish Sign Language Corpus (Version 8). Department of Linguistics, Stockholm University. URL.

Svenskt teckenspråkslexikon. (2018). Swedish Sign Language Dictionary online. Department of Linguistics, Stockholm University. teckensprakslexikon.su.se.

Recognizing Non-manual Signals in Filipino Sign Language

Joanna Pauline Rivera, Clement Ong
De La Salle University
Taft Avenue, Manila 1004 Philippines
joanna_rivera@dlsu.edu.ph, clem.ong@delasalle.ph

Abstract

Filipino Sign Language (FSL) is a multi-modal language that is composed of manual signlas and non-manual signals. Very minimal research is done regarding non-manual signals (Martinez and Cabalfin, 2008) despite the fact that non-manual signals play a significant role in conversations as it can be mixed freely with manual signals (Cabalfin et al., 2012). For other Sign Languages, there have been numerous researches regarding non-manual; however, most of these focused on the semantic and/or lexical functions only. Research on facial expressions in sign language that convey emotions or feelings and degrees of adjectives is very minimal. In this research, an analysis and recognition of non-manual signals in Filipino Sign Language are performed. The non-manual signals included are Types of Sentences (i.e. Statement, Question, Exclamation), Degrees of Adjectives (i.e. Absence, Presence, High Presence), and Emotions (i.e. Happy, Sad, Fast-approaching danger, Stationary danger). The corpus was built with the help of the FSL Deaf Professors, and the 5 Deaf participants who signed 5 sentences for each of the types in front of Microsoft Kinect sensor. Genetic Algorithm is applied for the feature selection, while Artificial Neural Network and Support Vector Machine is applied for classification.

Keywords: non-manual signal, Filipino Sign Language, Kinect, machine learning

1. Introduction

Filipino Sign Language (FSL) is a communication medium among the Deaf in the Philippines. It originally rooted from American Sign Language (ASL) (Hurlbut, 2008) but soon developed and became a separate language from ASL as Filipino Deaf used it in communication at school passing Filipino signs that emerged naturally through generations and adding new signs, especially those that are related to technology (Apurado and Agravante, 2006).

FSL has five components: hand shape, location, palm orientation, movement, and non-manual signal (Philippine Deaf Research Center and Philippine Federation of the Deaf, 2004).

1. **Hand shape** is the arrangement of the fingers and their joints.
2. **Location** is the place where the hand/s is/are.
3. **Palm orientation** is where the palm is facing.
4. **Movement** is the change in hand shape and/or path of the hands.
5. **Non-manual signals** are facial expressions and/or movement of the other parts of the body that goes with the signing.

Several researches in the computing field regarding FSL have been conducted despite the only recent linguistic researches about it. However, most studies in Filipino Sign Language (FSL) focuses on recognizing manual signals only. Very minimal research was done regarding non-manual signals and its integration with the manual signals (Martinez and Cabalfin, 2008) even though non-manual signals play a significant role in conversations as it can be mixed freely with manual signals (Cabalfin et al., 2012).

A common misconception in Sign Language Recognition Systems is approaching the problem through Gesture Recognition (GR) alone (Cooper et al., 2011). Sign language is a multi-modal language that has two components: manual and non-manual signals. Manual signals are hand gestures, positions and shapes which convey lexical information. Non-manual signals are facial expressions, head movements, and upper body posture and movements which express syntactic and semantic information. These signals usually co-occur with manual signals often changing its meaning (Nguyen and Ranganath, 2008). Hence, SLR systems or techniques would not yield effective results without the non-manual signals (Von Agris et al., 2008).

As researchers realize the significance of non-manual signals, some studies focusing on facial expression recognition in Sign Languages were conducted (Von Agris et al., 2008). These studies focus more on Grammatical Facial Expressions (GFE) that convey semantic functions such as WH-question, Topic and Assertion.

Some of the studies that focus on recognizing facial expressions on Sign Language include: Grammatical Facial Expressions Recognition with Machine Learning (de Almeida Freitas et al., 2014), Recognition of Non-manual Markers in American Sign Language (Metaxas et al., 2012), and Spatial and Temporal Pyramids for Grammatical Expression Recognition of American Sign Language (Michael et al., 2009). Most of these studies differ in the data representation, and the machine learning technique. One similarity among them is that they all focused on GFEs that convey semantic functions such as WH-question, Topic and Assertion. On the other hand, facial expressions in sign languages are not limited to those semantic functions. For instance, facial expressions in ASL are used to convey degrees of adjectives (e.g. color intensity), adverbial information (e.g. carelessly) and emotions as well (Martinez and Cabalfin, 2008).

Similarly, non-manual signals in FSL is used to convey lexical information, types of sentences (what is referred to as semantic information in other studies), degrees of adjectives, and emotions.

This study aims to recognize and analyze non-manual signals in FSL using the Microsoft Kinect and Machine Learning. The rest of the paper is organized as follows. In section 2, the process of building the corpus with the help of FSL signers is described. In section 3, the feature extraction method is discussed. In section 4, the Machine Learning techniques applied are enumerated. In section 5, the results and analysis for each non-manual signal category are explained. Lastly, the research is concluded and the recommendations are listed in section 6.

2. Corpus Building

2.1. Interviews with FSL Deaf Professors

To understand more about FSL, an interview was conducted with Ms. Maria Elena Lozada, a Deaf professor from the School of Deaf Education and Applied Studies at De La Salle - College of Saint Benilde. According to her, some non-manual signals in FSL are part of its lexicon such as "thin" and "fat". Others are used to differentiate sentences with the same signs but different semantics. For instance, the statement "John likes Mary" is signed the same way as the question "Does John like Mary?". The two sentences are differentiated using different non-manual signals. There are also non-manual signals that are used to convey the degrees of adjectives (e.g. "angry" and "very angry"). Lastly, as Filipinos were born naturally expressive, non-manual signals are mostly used to convey emotions or feelings (Lozada, 2016).

Another interview was conducted with Mr. Rey Alfred Lee, a Deaf professor from Filipino Sign Language Learning Program, Center Education Assessment for the Deaf, and School of Deaf Education and Applied Studies at De La Salle - College of Saint Benilde. From the interview, it was concluded that FSL uses Affective Facial Expressions (AFE) and Grammatical Facial Expressions (GFE), similar to other Sign Languages. AFEs in FSL are used to show emotions and feelings, while GFEs are used to convey lexical information, types of sentences, and degrees of adjectives. The types of sentences are further subdivided into question, statement, and exclamation, while the basic degrees of adjectives are subdivided into absence, presence, and high presence (Lee, 2016). In other researches, it is stated that GFE differ from AFE in terms of the facial muscles used (McCullough and Emmorey, 2009) and its behavior, such as form and duration (Muller, 2014).

Both professors are asked about the importance of non-manual signals in FSL. According to Ms. Lozada, although it is possible to use FSL without facial expressions, it would be difficult to carry out a conversation especially when telling a story which involves facial expressions to convey types of sentences, emotions, and degrees of adjectives (Lozada, 2016). According to Mr. Lee, being able to recognize the degrees of adjectives, specifically feelings, and emotions can also help the medical doctors and psychologists in determining the amount of pain that the signer feels and the emotional state of the patient (Lee, 2016).

In line with this, this research focuses on non-manual signals in FSL that convey Types of Sentences (i.e. statement, question, and exclamation), Degrees of Adjectives (i.e. absence, presence, and high presence), and Emotions.

The emotions considered are the four basic emotions (i.e. happy, sad, fast-approaching danger and stationary danger) (Jack et al., 2014). In their work, they have shown that there are only four basic emotions which were only discriminated into six (i.e. happy, sad, fear, surprise, disgust and anger) as time passed by. Fear and surprise can be combined as they both belong in the fast-approaching danger, while disgust and anger both belong in the stationary danger.

Table 1 shows a summary of the types of Facial Expressions in FSL that were used for this research. These types were used as labels by the FSL annotator.

<table>
<tr><td rowspan="4">Affective Facial Expressions</td><td rowspan="4">Emotions</td><td>Happy</td></tr>
<tr><td>Sad</td></tr>
<tr><td>Stationary danger</td></tr>
<tr><td>Fast-approaching danger</td></tr>
<tr><td rowspan="6">Grammatical Facial Expressions</td><td rowspan="3">Types of Sentences</td><td>Statement</td></tr>
<tr><td>Question</td></tr>
<tr><td>Exclamation</td></tr>
<tr><td rowspan="3">Degrees of Adjectives</td><td>Absence</td></tr>
<tr><td>Presence</td></tr>
<tr><td>High Presence</td></tr>
</table>

Table 1: Categories of Facial Expressions in Filipino Sign Language

2.2. Data Collection with FSL Signers

There are already existing corpus for FSL. However, these are built to focus on the manual signals data. The different types of non-manual signals may not be represented well on these corpus.

Thus, data collection is performed using Microsoft Kinect for Windows v2.0 sensor (Kinect sensor) (Microsoft, 2016). The Kinect sensor has a depth sensor, a color camera, an infrared (IR) emitter, and a microphone array that allow tracking of the location, movement, and voice of people (Zhang, 2012).

The 3D videos are collected from 5 participants, 20-24 year old third-year FSL students. Two of them are male while three of them are female. Their learning and actual experience in FSL is approximately 2-4 years. With regards to facial expressions in FSL, most of them have learned it in school about 1-3 months, while some of them have been using it for 1-3 years.

5 sentences for each type of facial expression, a total of 50 sentences, were signed by each participant. To assure that all samples are appropriate for its corresponding type, these sentences were formulated with the guidance of the FSL Deaf professor of the participants, Ms. Lozada. Refer to Table 2 for the complete list of sentences used for this study.

Type	Sentences
Happy	1. Thank you. 2. The trip is exciting. 3. The show is amazing. 4. I am proud of you! 5. Our team won!
Stationary Danger	1. I hate you! 2. You are disgusting! 3. I don't like you! 4. You are so slow! 5. Stay away from me!
Fast-approaching Danger	1. I am scared. 2. I am nervous. 3. I am worried. 4. I saw a ghost. 5. I am shocked!
Sad	1. I am sorry. 2. My dog died. 3. I am alone. 4. I am heartbroken. 5. I failed the exam.
Question	1. Does John like Mary? 2. Are you sick? 3. Is it new year? 4. How are you? 5. How old are you?
Statement	1. John likes Mary. 2. You are sick. 3. It is new year. 4. I am fine. 5. I am 12 years old.
Exclamation	1. John likes Mary! 2. You are sick! 3. Happy new year! 4. Good morning! 5. Good noon!
Absence	1. My head is not painful. 2. I do not like you. 3. I am not tired. 4. You are not slow. 5. This is not hard.
Presence	1. My head is painful. 2. I like you. 3. I am tired. 4. You are slow. 5. This is hard.
High Presence	1. My head is very painful. 2. I like you very much. 3. I am so tired. 4. You are so slow. 5. This is very hard.

Table 2: Sample Sentences for each of the types of Non-manual Signals in FSL

2.3. Data Annotation with FSL expert

Supposedly, the annotation label of each sentence is their intended type since the facial expressions are acted, see Table 2 for the intended type for each of the sentences. However, initial experiments show very poor performances reaching the highest accuracy of 26% using Artificial Neural Network (ANN). Looking at the confusion matrix shown in Table 3, it can be deduced that the classes are very confused with each other, meaning there are similarities between them.

true=	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**
pred. a	**5**	4	5	2	3	1	6	1	0	0
pred. b	4	**4**	3	2	1	1	0	2	1	0
pred. c	4	5	**7**	0	3	2	1	2	3	1
pred. d	0	2	1	**7**	3	4	0	2	4	2
pred. e	1	0	1	2	**4**	6	4	1	1	4
pred. f	3	1	2	2	4	**2**	1	2	0	2
pred. g	1	1	1	4	2	1	**5**	2	1	3
pred. h	5	5	2	1	2	3	3	**10**	1	2
pred. i	1	3	3	3	1	1	2	1	**12**	2
pred. j	1	0	0	2	2	4	3	2	2	**9**

Table 3: Confusion matrix during initial experiment using ANN where: a=question, b=statement, c=Exclamation, d=absence, e=presence, f=high presence, g=stationary, h=fast, i=happy, and j=sad

With a consultation with an FSL annotation expert, co-occurrences of the different classes in a sample are discovered. As a result, there is a maximum of three labels in an instance. For example, the facial expression for specific sign/s can be a combination of question (one of the types of sentences), presence (one of the degrees of adjectives), and sad (one of the emotions). Thus, individual experiments were conducted for Types of Sentences, Degrees of Adjectives and Emotions, each applying the classification techniques.

3. Feature Extraction

Color images, depth images, audio input, and skeletal data from Kinect sensor are processed with the help of Microsoft Kinect for Windows Software Development Kit 2.0 (Kinect SDK) (Microsoft, 2016) to extract the features.

The face orientation, Shape Units (SU), and Animation Units (AU) are used as features for this study as most international research works have concluded that the eyes, eyebrows, mouth, nose, and head pose must be well-represented to achieve effective recognition. The face orientation is the computed center of the head which is used to calculate the angle rotations of the head with respect to the optical center of the camera of Kinect sensor (i.e. pitch, yaw, and roll). The SUs are the weights that indicate the differences between the shape of the face tracked and the average shape of a person which is derived using the Active Appearance Models (AMM) of (Smolyanskiy et al., 2014). These SUs are used to indicate the neutral shape of the face which is derived from the first few frames. The AUs are the deltas in the facial features of the face tracked from the neutral shape.

In summary, the 20 features used are the pitch, yaw and roll angles, and the seventeen AUs shown in Table 4. Most of the values range from 0 to 1. Negative minimum value indicates delta on the opposite direction. For example, if the delta for EyebrowLowerer is -1, the eyebrows are raised instead of lowered.

Movement	**Min Value**	**Max Value**
Pitch		
Yaw		
Roll		
JawOpen	0	+1
LipPucker	0	+1
JawSlideRight	-1	+1
LipStretcherRight	0	+1
LipStretcherLeft	0	+1
LipCornerPullerRight	0	+1
LipCornerPullerLeft	0	+1
LipCornerDepressorLeft	0	+1
LipCornerDepressorRight	0	+1
LeftCheekPuff	0	+1
RightCheekPuff	0	+1
LeftEyeClosed	0	+1
RightEyeClosed	0	+1
RighteyebrowLowerer	-1	+1
LefteyebrowLowerer	-1	+1
LowerlipDepressorLeft	0	+1
LowerlipDepressorRight	0	+1

Table 4: Face Orientation and Animation Units with the minimum and maximum weights

4. Machine Learning

Before the data has undergone classification, some preprocessing tasks are performed. Particularly, only samples based from peak facial expressions are selected since some frames between the neutral and peak facial expressions showed hand occlusions on the face. This also ensures that rising and falling facial expressions are excluded from the samples. Afterwards, uniform undersampling is applied since the data for Degrees of Adjectives is imbalanced. Normalization through z-transformation is also applied due to the different ranges of feature values.

Then, feature selection is applied to determine the most effective features for each category. The Wrapper Subset Evaluation dominated in terms of improving the accuracy; however, it is computationally expensive (Hall and Holmes, 2003). Thus, Genetic Algorithm is applied to reduce the amount of resources needed.

Some of the most commonly used classifiers in recent studies regarding Facial Expression Recognition in Sign Language are Artificial Neural Network (ANN), and Support Vector Machine (SVM) (Mao and Xue, 2011). ANN with a learning rate of 0.3, and SVM with a kernel type of radial basis function are applied for this study. Then, the validation technique used is k-fold Cross-Validation while the performance metrics are accuracy and kappa.

5. Experiments, Results and Analysis

Several experiments are conducted to analyze the types of non-manual signals in FSL. These experiments can be categorized into 3: Participants-based, Features-based, and Class-based.

The Participants-based experiments are subdivided into Partitioning by Participants and Partitioning by Sentences. In Partitioning by Participants setup, there are five folds for the validation phase. In each fold, there are four participants in the training set while there is one participant in the test set. In Partitioning by Sentences, there are a total of 10 folds for the validation phase. In each fold, 90% of the sentences are in the training set, while 10% are in the test set.

Using Participants-based experiments, findings indicate that there are not much differences on the performances for all categories between the Participant-based experiments as shown in Figure 1 and Figure 2. This suggests that introducing a new face would not have much impact on the classification. This is because AUs are deltas and not the exact points on the face. Thus, different facial structures would not matter that much as long as the participants are all expressive. In Sign Language, facial expressions are significant; thus, signers are usually expressive.

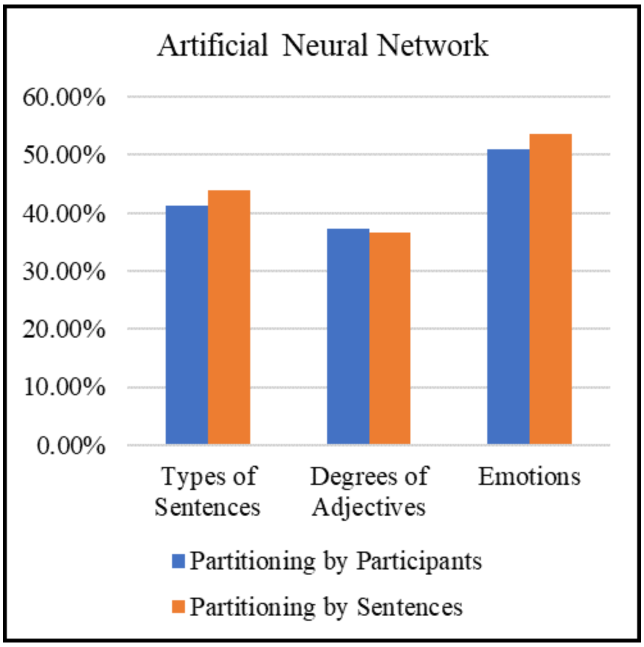


Figure 1: Comparison of performances using Partitioning-based setups for ANN

The Features-based experiments rely on adding classes from other categories as features. For example, the features for Degrees of Adjectives may include Fast-approaching danger, Stationary danger, Happy, and Sad. This is an attempt to represent the intensities for degrees of adjectives, and the co-occurrences of the different categories in one in-

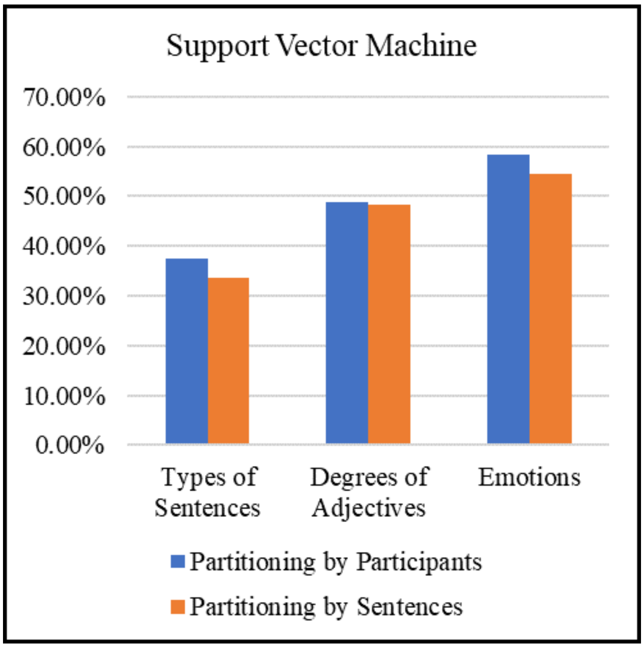


Figure 2: Comparison of performances using Partitioning-based setups for SVM

stance. The idea here is that the Degree of Adjective is Presence if the feature values are within the average given a co-occurring Emotion. It is High Presence if the feature values are higher than the average given a co-occurring Emotion.

Using Features-based experiments, findings indicate that adding classes from other categories as features are effective in representing the intensities, and the co-occurrences of the different categories in one instance reaching an increase of 17% to 30% recognition rate.

The Class-based experiments are subdivided into: One versus Many and One Class Removed. In One versus Many, one class is retained while the remaining classes are merged to form a new class. For example, the possible classes for Degrees of Adjectives are Presence or Not, Absence or Not, and High Presence or Not. In One Class Removed, one class is not included for each category during the experiments. For examples, the possible classes for Types of Sentences are Statement or Question, Statement or Exclamation, and Question or Exclamation.

Using Class-based experiments, highest performances for all categories are achieved which implies that some essential features for other classes are not represented. This is because motions and context are not represented which are significant for some of the classes based on the direct observation of the video data.

5.1. Types of Sentences

Distinction of Question and a confusion between Statement and Exclamation are observed in class-based experiments as shown in Table 5. It is also observed that Question has more distinction with Statement than with Exclamation.

Similar findings and further improvements were observed using Feature-based setups. Adding Emotions and Degrees of Adjectives as features resulted to the highest performances. Hence, Types of Sentences are highly affected by the co-occurrences of the other categories. Refer to Table 6 for the performances in accuracies and kappas.

Classes	ANN		SVM	
	Acc.	Kappa	Acc.	Kappa
Exclamation or Not	55.00%	0.220	53.33%	0.110
Question or Not	61.33%	0.231	62.33%	0.264
Statement or Not	63.33%	0.267	53.33%	0.067
Statement-Question	57.67%	0.199	60.00%	0.265
Statement-Exclamation	46.67%	-0.060	43.33%	-0.070
Question-Exclamation	60.00%	0.210	46.67%	-0.040

Table 5: Comparison of the performances of ANN, and SVM using Class-based setups for Types of Sentences

Setup	ANN		SVM	
	Acc.	Kappa	Acc.	Kappa
Emotions and Degrees	60.00%	0.409	58.00%	0.363
Emotions	40.00%	0.089	30.50%	-0.008
Degrees	43.00%	0.114	52.00%	0.277

Table 6: Comparison of the performances of ANN, and SVM using Feature-based setups for Types of Sentences

Since adding Emotions and Degrees of Adjectives as features lead to higher performances, Genetic Algorithm for feature selection is applied after performing this setup. Refer to Table 7 for the list of features selected along with the weights. Applying feature selection and adding Emotions and Degrees of Adjectives resulted to performances reaching accuracy of 76.00% and kappa of 0.619 using ANN. Similar behavior on the confusion matrices are also observed but the distinction between Statement and Exclamation improved.

From the observation of the videos, Question is mostly characterized by eyes, eyebrows, lip, and head movements which makes it distinct from the other classes. The eyes movements are captured by LeftEyeClosed and RightEyeClosed. The eyebrows movements are captured by RightEyebrowLowerer and LeftEyebrowLowerer. The lip movements (i.e. lip corners pulled downwards) are captured by LipCornerPullerLeft, LipCornerPullerRight, LipCornerDepressorRight, and LipCornerDepressorLeft. Lastly, the head movements are captured by pitch, yaw, and roll. Since all the distinct characteristics of Question are represented by the head rotation angles and seventeen AUs, Question always has the highest precisions and recalls.

On the other hand, Statement and Exclamation are always confused. Looking at the videos, these classes do not have much distinguishing features. Statement is like a neutral facial expression that changes based on the current Degree of Adjective or Emotion. When it is mixed with the other classes, it becomes similar to Exclamation.

Feature	Weight
emotions = n	1
emotions = happy	1
emotions = stationary	1
emotions = fast	1
degrees = absence	1
degrees = n	1
pitch	1
yaw	1
JawSlideRight	1
LipCornerPullerRight	1
LipCornerDepressorRight	1
RightcheekPuff	1
LefteyeClosed	1
RighteyebrowLowerer	1
LowerlipDepressorLeft	1

Table 7: Features Selected using Genetic Algorithm on Types of Sentences

Aside from this, only the difference in speed of the motions are observed. In this study, only the peak facial expressions are selected so the motions are not captured. Also, the head rotation angles and seventeen AUs alone cannot handle motions since these features are only concerned with the deltas between the current and the neutral facial expression.

5.2. Degrees of Adjectives

Distinction between Absence and High Presence is observed on One Class Removed of Class-based experiments as shown in Table 8. On the other hand, Presence cannot be distinguished from the rest of the classes as shown in One versus Many of Class-based experiments.

Setup	ANN		SVM	
	Acc.	Kappa	Acc.	Kappa
Presence or not	42.44%	-0.146	40.32%	-0.160
Absence or not	70.24%	0.372	70.00%	0.371
High Presence or not	59.09%	0.174	68.11%	0.354
High Presence-Absence	82.38%	0.631	87.14%	0.727
Presence-Absence	50.48%	0.019	52.86%	0.078
Presence-High Presence	46.67%	-0.059	56.97%	0.164

Table 8: Comparison of the performances of ANN, and SVM using Class-based setups for Degrees of Adjectives

Absence and High Presence are like polar opposites which is why they can easily be distinguished from each other. On the other hand, Presence is like a neutral facial expression similar to Statement. It becomes similar to the other classes when mixed with Emotions.

High Presence can be differentiated from Presence based on the intensity of the facial expression of the sentence. The intensity is represented through adding other classes as features. Results shown in Table 9 indicate that adding Emotions as features yield the best performances among Features-based experiments. This validates the idea that the Degree of Adjective is Presence if the feature values are within the average given a co-occurring Emotion. It is High Presence if the feature values are higher than the average given a co-occurring Emotion.

Setup	ANN		SVM	
	Acc.	Kappa	Acc.	Kappa
Emotions and Sentences	47.22%	0.218	54.78%	0.330
Emotions	51.22%	0.265	62.33%	0.435
Sentences	33.67%	0.014	48.56%	0.231

Table 9: Comparison of the performances of ANN, and SVM using Feature-based setups for Degrees of Adjectives

Absence can be differentiated from Presence by detecting the motion of head shaking. However, the head rotation angles and seventeen AUs only represent the delta between the neutral face and the peak facial expression. The motion yawing to the left or right can be captured by the AUs, but not the whole motion of head shake.

Genetic Algorithm is applied for feature selection after adding Emotions as features. Refer to Table 10 for the complete list with the weights. Without removing classes or merging classes, the highest accuracy reached 70.89% with a kappa of 0.562 using SVM.

Feature	Weight
yaw	1
JawOpen	1
LipStretcherRight	1
LipCornerPullerLeft	1
RighteyebrowLowerer	0.938899
LefteyeClosed	0.89875
LipCornerDepressorLeft	0.735377
LowerlipDepressorLeft	0.429755
LeftcheekPuff	0.413458
pitch	0.221545

Table 10: Features Selected using Genetic Algorithm on Degrees of Adjectives

5.3. Emotions

Results from Class-based experiments indicate that Happy and Fast-approaching danger are distinct from the other classes using One versus Many (i.e. Happy or Not, and Fast-approaching danger or Not), while Sad and Stationary danger are confused with each other as shown using One Class Removed (i.e. Happy or Sad or Stationary danger, and Fast-approaching danger or Sad or Stationary danger). Refer to Table 11 for the results of Class-based experiments.

In contrary to the effect of Features-based experiments on Types of Sentences and Degrees of Adjectives, adding classes from other categories as features did not have good effect on the performances. Refer to Table 12 for the results of Features-based experiments. It is shown that the highest

Setup	ANN		SVM	
	Acc.	Kappa	Acc.	Kappa
Fast or not	68.50%	0.367	73.50%	0.467
Happy or not	87.00%	0.739	85.33%	0.718
Sad or not	65.67%	0.291	74.67%	0.495
Stationary or not	74.00%	0.470	70.00%	0.398
Happy-Sad-Stationary	59.72%	0.398	55.00%	0.332
Fast-Sad-Stationary	55.71%	0.340	51.61%	0.278
Fast-Happy-Stationary	68.57%	0.529	79.46%	0.687
Fast-Happy-Sad	76.61%	0.642	78.93%	0.671

Table 11: Comparison of the performances of ANN, and SVM using Class-based setups for Emotions

accuracy of 67% is achieved when Sentences and Degrees are added as features. However, this is lower than when only Genetic Algorithm is applied resulting to 72.91% accuracy. This implies that Emotions is not affected by the co-occurrence of the other categories.

Setup	ANN		SVM	
	Acc.	Kappa	Acc.	Kappa
Sentences and Degrees	59.27%	0.454	67.00%	0.556
Degrees	54.73%	0.393	58.18%	0.443
Sentences	62.36%	0.492	65.00%	0.529

Table 12: Comparison of the performances of ANN, and SVM using Feature-based setups for Emotions

Happy and Fast-approaching danger have characteristics that make them distinct from the other classes. Happy is mostly characterized by a smiling face, while Fast-approaching danger are mostly characterized by eyes and/or mouth wide opened. On the other hand, the characteristics of Sad and Stationary danger are very similar which makes it difficult for the classifier to distinguish between the two. Stationary danger and sad are mostly characterized by a frowning face. A possible reason why the annotators can recognize it is their knowledge about the context shown by the gestures.

5.4. Summary

Without removing or merging classes, highest performances are achieved by adding classes from other categories as features and/or applying genetic algorithm for feature selection. For Types of Sentences, Emotions and Degrees of Adjectives are added as features and genetic algorithm is applied, reaching the highest accuracy of 76.00% and kappa of 0.619 using ANN. For Degrees of Adjectives, Emotions are added as features and genetic algorithm is applied, reaching the highest accuracy reached 70.89% with a kappa of 0.562 using SVM. For Emotions, genetic algorithm is applied reaching the highest performance of 72.91% accuracy and 0.639 kappa.

6. Conclusion and Recommendations

In this study, the different non-manual signals, specifically Types of Sentences, Degrees of Adjectives, and Emotions are recognized and analyzed towards aiding the communication between the Deaf community, and the medical doctors, psychologists, and other non-signers.

Based on the experiments conducted, AUs are effective in representing different facial structures of the signers, but motions, intensities, and co-occurrences of classes from other categories must also be well-represented. Representing the intensities and co-occurrences by adding classes from other categories as features yielded better performances. However, confusion matrices show that the representation of intensities must still be improved. In addition, the gesture data is important as it shows the context which can further help in distinguishing the facial expressions. As stated by the FSL experts and annotators, knowing the meaning of gestures help them annotate the facial expressions. Without the seeing the gestures it would be difficult for them to distinguish the facial expressions.

In line with the conclusion, in addition to head rotation angles and AUs, motions must be captured to represent the data better. In the studies of (Von Agris et al., 2008), (de Almeida Freitas et al., 2014), and (Nguyen and Ranganath, 2010), representations of motions through the inclusion of temporal information such as Sliding Window and Spatio-Temporal Pyramids improved their recognition rates. In the studies of (Metaxas et al., 2012), and (Nguyen and Ranganath, 2010), machine learning techniques that can handle temporal dynamics by making use of sequential data were applied such as Hidden Markov Model and Conditional Random Field respectively. Motions are not represented in this study since some frames were dropped due to hand occlusions. Based from the other works, removing the frames with hand occlusions is not the solution to the problem. Feature detection techniques that can handle occlusions must be applied such as Kanade-Lucas-Tomasi (KLT) with Bayesian Feedback Mechanism and Non-parametric Adaptive 2D-3D Tracking in the studies of (Metaxas et al., 2012) and (Nguyen and Ranganath, 2010) respectively, instead of AAM-based methods.

Aside from motions, intensities and co-occurrences must also be represented well. In this study, an attempt to represent these is adding classes from other categories as features. Significantly better performances were observed using this setup. However, this approach can still be improved. One way to recognize the intensities could be applying Convolutional Neural Network (CNN) for classification which is one of the state-of-the-art approach of deep learning for images.

Lastly, an integration of the gesture data can be explored as it contains the context that might be significant in distinguishing the facial expressions. As the annotator and other FSL experts have mentioned, annotating the data without seeing the gestures is possible but it would be difficult.

7. Acknowledgements

The authors would like to thank the anonymous participants from the De La Salle-College of St. Benilde who agreed to participate in the data gathering for the corpus building.

The authors would also like to express their deepest gratitude to Ms. Maria Elena Lozada for her support in the data collection, and Mr. Rey Alfred Lee and Mr. John Xander Baliza for imparting their knowledge on Filipino Sign Language.

8. Bibliographical References

Apurado, Y. S. and Agravante, R. L. (2006). The phonology and regional variation of Filipino Sign Language: considerations for language policy. In *9th Philippine Linguistic Congress*.

Cabalfin, E. P., Martinez, L. B., Guevara, R. C. L., and Naval, P. C. (2012). Filipino Sign Language recognition using manifold projection learning. In *TENCON 2012-2012 IEEE Region 10 Conference*, pages 1–5. IEEE.

Cooper, H., Holt, B., and Bowden, R. (2011). Sign language recognition. In *Visual Analysis of Humans*, pages 539–562. Springer.

de Almeida Freitas, F., Peres, S. M., de Moraes Lima, C. A., and Barbosa, F. V. (2014). Grammatical facial expressions recognition with machine learning. In *FLAIRS Conference*.

Hall, M. A. and Holmes, G. (2003). Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. *IEEE Trans. on Knowl. and Data Eng.*, 15(6):1437–1447, November.

Hurlbut, H. M. (2008). Philippine signed languages survey: a rapid appraisal. *SIL Electronic Survey Reports*, 10:16.

Jack, R. E., Garrod, O. G., and Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current biology*, 24(2):187–192.

Lee, R. A. (2016). Facial expressions in Filipino Sign Language, March. Personal Communication.

Lozada, M. E. (2016). Facial expressions in Filipino Sign Language, February. Personal Communication.

Mao, X. and Xue, Y. (2011). Human Emotional Interaction.

Martinez, L. and Cabalfin, E. P. (2008). Sign language and computing in a developing country: a research roadmap for the next two decades in the Philippines. In *PACLIC*, pages 438–444.

McCullough, S. and Emmorey, K. (2009). Categorical perception of affective and linguistic facial expressions. *Cognition*, 110(2):208–221, February.

Metaxas, D. N., Liu, B., Yang, F., Yang, P., Michael, N., and Neidle, C. (2012). Recognition of nonmanual markers in American Sign Language (ASL) Using nonparametric adaptive 2d-3d face tracking. In *LREC*, pages 2414–2420. Citeseer.

Michael, N., Metaxas, D., and Neidle, C. (2009). Spatial and temporal pyramids for grammatical expression recognition of American Sign Language. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, pages 75–82. ACM.

Microsoft. (2016). Kinect hardware requirements and sensor setup.

Muller, C. (2014). *Body - Language - Communication*. Walter de Gruyter GmbH & Co KG, October. Google-Books-ID: QAGTBgAAQBAJ.

Nguyen, T. D. and Ranganath, S. (2008). Tracking facial features under occlusions and recognizing facial expressions in sign language. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–7. IEEE.

Nguyen, T. D. and Ranganath, S. (2010). Recognizing Continuous Grammatical Marker Facial Gestures in Sign Language Video. In *Computer Vision - ACCV 2010*, pages 665–676. Springer, Berlin, Heidelberg, November.

Philippine Deaf Research Center and Philippine Federation of the Deaf. (2004). *An Introduction to Filipino Sign Language: (a). Understanding structure*. An Introduction to Filipino Sign Language. Philippine Deaf Resource Center, Incorporated.

Smolyanskiy, N., Huitema, C., Liang, L., and Anderson, S. E. (2014). Real-time 3d face tracking based on active appearance model constrained by depth data. *Image and Vision Computing*, 32(11):860–869, November.

Von Agris, U., Knorr, M., and Kraiss, K.-F. (2008). The significance of facial features for automatic sign language recognition. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE.

Zhang, Z. (2012). Microsoft Kinect Sensor and Its Effect. *IEEE MultiMedia*, 19, April.

The Hong Kong Sign Language Browser

Felix Y.B. Sze, Kloris T.L. Lau, Kevin M.H. Yu
Centre for Sign Linguistics and Deaf Studies, The Chinese University of Hong Kong
G01, Academic Building Two, The Chinese University of Hong Kong, Shatin, Hong Kong
felix_csls@cuhk.edu.hk, kloris_cslds@cuhk.edu.hk, kevin_cslds@cuhk.edu.hk

Abstract

This paper describes the design of the Hong Kong Sign Language Browser which was established for providing accessible online resources on the lexical variations of HKSL in order to support the promotion of sign language and other sign-related services in the local community. With continuous funding support from the government since 2012, local Deaf organizations and Deaf signers of diverse backgrounds are invited to contribute their sign language knowledge in the data collection and evaluation process. Each year Deaf informants proficient in HKSL are invited to CSLDS to provide signing data to a pre-defined list of lexical targets. Their signing data are analyzed and variants are identified. These video data are then placed in an online platform for local Deaf organizations for rating and comments, and they can contribute data as well if there are additional variants not yet covered in the initial round of data collection. Once finalized, the lexical variants are placed in the Hong Kong Sign Language Browser for free public access. For each lexical target, each variant is indicated by a different color. Variants that are more commonly used and seen by Deaf organizations are listed first whereas the least common variants are listed last.

Keywords: sign language browser development, Hong Kong Sign Language, Deaf community involvement

1. Introduction

The Hong Kong Sign Language Browser (HKSL Browser) is an online database that aims at documenting lexical variants in HKSL for community use. It was established by the Centre for Sign Linguistics and Deaf Studies (CSLDS) of the Department of Linguistics and Modern Languages at The Chinese University of Hong Kong (CUHK) in 2012 with continuous funding support from the Labor and Welfare Bureau of the Hong Kong Special Administration Region Government. At present, there are 1,524 lexical targets with a total of 4,544 variants.

2. Objectives of Browser

The major objective of establishing the HKSL Browser is to provide publicly accessible online lexical resources for promoting HKSL in the local community and supporting the development of sign-related services such as sign interpretation and Deaf education. As a natural language, HKSL exhibits a considerable degree of lexical variations across signers, and these variants stem from differences of the signers' age and educational backgrounds, or are the natural outcomes of phonological processes (Sze, Lo, Lo & Chu 2012, 2013). Although variation is a normal linguistic phenomenon, many Deaf people in Hong Kong are worried about the existence of variations, which led to unnecessary conflicts among themselves from time to time. Hence, educating the Deaf community about the need to respect and document variation is another objective of the HKSL Browser. To ensure a wide coverage of variations in the Browser, local Deaf organizations and Deaf signers of diverse backgrounds are invited to take part in the data collection and evaluation process.

Although the HKSL Browser aims at documenting lexical variations, its design and data collection procedure are not intended to facilitate vigorous sociolinguistic research. In a typical sign language variation corpus that is research-oriented, selection of informants and data types are carefully controlled to ensure an adequate coverage of social factors that may possibly underlie linguistic variation. In the British Sign Language Corpus (Schembri, Fenlon, Rentelis, Reynolds & Cormier 2013), for instance, the Deaf informants were recruited from 8 different cities across the United Kingdom, representing a balanced mix of men and women of different age ranges, family backgrounds (e.g., whether their parents were deaf and hearing), job types and ethnic groups. In addition, they all participated in the same data elicitation tasks including story-telling, free conversation, answering interview questions, and producing a limited set of lexical targets (102 key concepts). However, as our major objective is to produce online lexical resources for the general public for supporting sign language related services, we prioritize the wide coverage of lexical entries and variants over a strict control of common sociolinguistic variables of the signing informants in our data collection process. As we will point out later in this paper, we invite Deaf informants who are known to be fluent HKSL signers and are active members in the local community, and all the lexical variants placed in the Browser are confirmed by local Deaf organizations to be authentic, existing signs currently in use in the community. While the entries are useful for linguistic analysis at the lexical level, the background information offered by the Deaf organizations at least offer some preliminary data on the distributions of the variants which may benefit future sociolinguistic research.

3. Data Collection

The data collection procedure consists of two parts: initial elicitation and formal shooting. For each round of data collection, Deaf fluent signers of different age ranges and different education backgrounds are invited to an interview during which our deaf researchers elicit the target signs from them through either pictures (e.g., lexical targets that can be pictorially represented) or group discussion (e.g., legal or medical concepts that require prior explanation and clarification before data elicitation). Ensuring the diverse backgrounds of the informants is of vital importance as we want to collect as many lexical variants as possible. The elicited data are then coded in ELAN and the Deaf informants will be invited to come to our research centre again for a formal shooting. For informants who do not want to be a signing model for privacy reason, their lexical variants will be demonstrated instead by one of our Deaf researchers in the formal shooting sessions. Efforts will be made to ensure that the signing replication is as exact as possible.

When ready, the formal video clips of these lexical variants will be uploaded to an internal online platform for the Deaf informants to check whether the signs are fine. Figure (1) is a screenshot of this Video Checking Platform. The informants can indicate if the video clips correctly illustrate their signing (particularly if their signs are demonstrated by our Deaf researchers rather than themselves being the signing model). They can indicate if the clips are correct or incorrect. They can also indicate if the signs are natural HKSL signs, or signed Chinese. Further comments can be added if necessary.

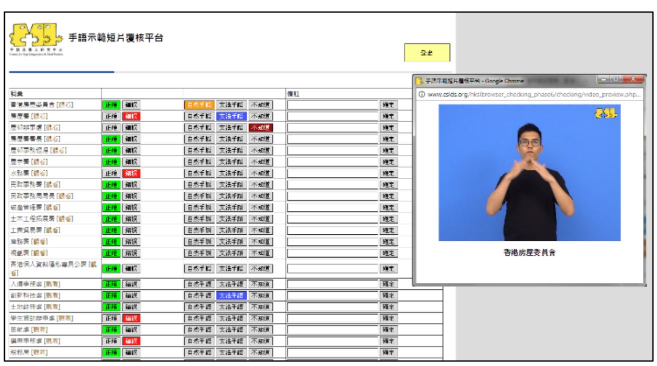


Figure 1: Screenshot of the Video Checking Platform

Clips that are unwanted by the Deaf informants will not be uploaded to the Browser. After the clips are checked, they will be further analyzed. The range of variants will be sorted out in our database, and their differences will be noted internally. The variants include both separate variants and phonological variants. Separate variants are those that are not phonologically related, and phonological variants share a certain degree of similarity in terms of the phonological parameters. This will be double-checked by another linguistically-trained researcher, with further discussion among several other Deaf colleagues if necessary.

The checked clips will then be uploaded to an evaluation platform called the Questionnaire Platform. For each lexical variant, there is a list of questions regarding its usage and distribution. Invitation is sent to all the Deaf organizations in Hong Kong, but participation is on a voluntary basis and a small amount of honorarium is offered as a token of gratitude for their participation. The participating organizations will be given a log-in account for completing the questionnaires. Their answers to the questions will be summarized and incorporated in the Browser for public viewing. Figure (2) is a screenshot of the first page of the questionnaire where all the variants are listed. To fill in the questionnaire for each sign, the informant needs to click the button on the right.

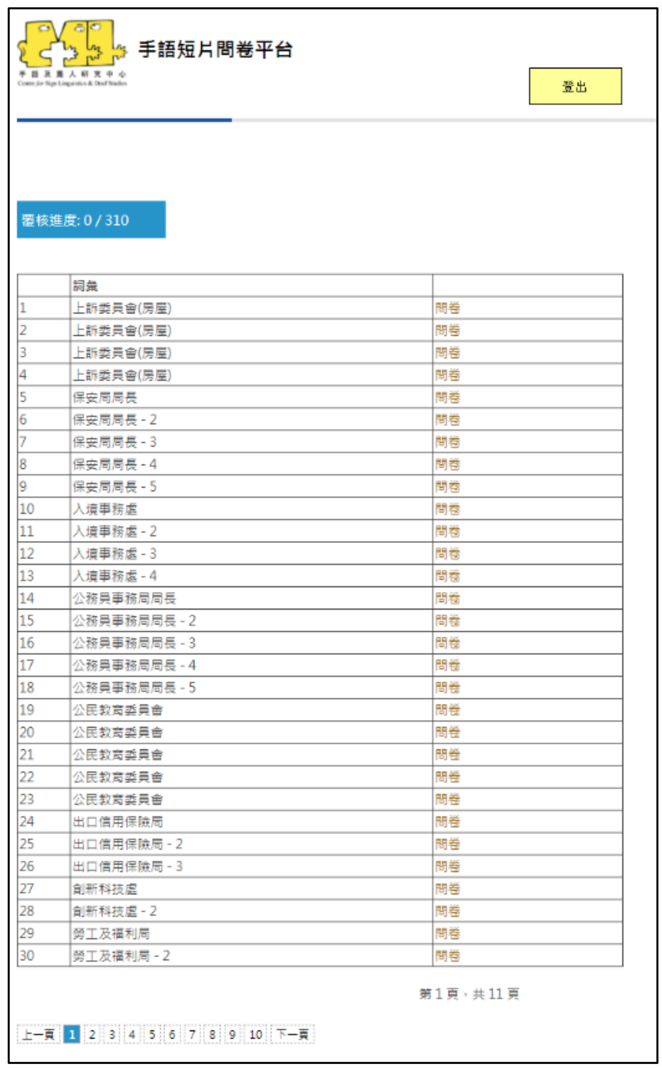


Figure 2: Screenshot of the Questionnaire Platform

These questions concern whether the informants have used or seen the signs before, whether they know the origin of the signs, and the background of signers who use these signs.

The Questionnaire Platform is an important component in the HKSL Browser as it collects feedback from different Deaf organizations to help us determine how common the variants are among local Deaf signers. Variants that are not used by any organizations and are seen by fewer than two organizations will not be listed in the Browser. The background information provided by the Deaf Organizations, e.g., whether the variant is a loan word from other sign languages, whether the variant is more commonly found among older or younger signers, etc., can help researchers and the general public to have a rough idea about the distribution of the variants. The signing preference of the participating Deaf organizations can be shown in the questionnaire results as well. Such information is found to be particularly useful to signing interpreters as they can now choose the appropriate variants when serving different parties. Besides, if the Deaf organizations notice that a variant has not yet been included in the Questionnaire Platform, they can upload a video clip as a suggestion and we will invite them to send a Deaf representative to do the formal shooting in the next round of data collection.

Figure (3) is a flow chart showing the sequence of questions on the Questionnaire Platform

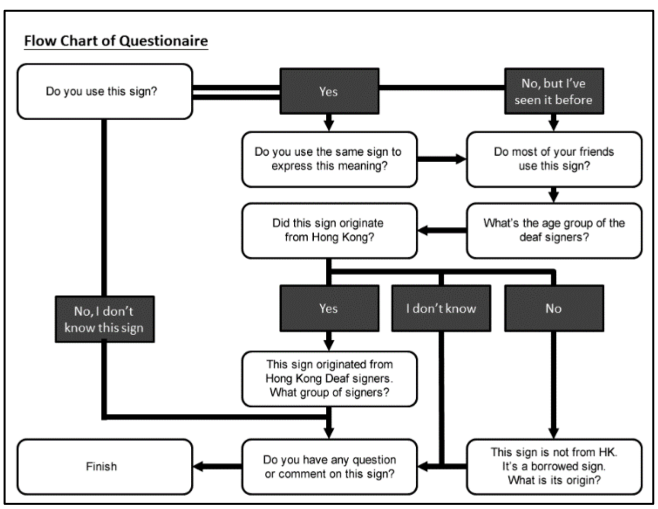


Figure 3: Flow chart of the questioning sequence on the Questionnaire Platform

Figure (4) is a screenshot showing the beginning of the Questionnaire. The video on the top is the target variant. The questions are presented in both written Chinese and HKSL on the left side of the platform. The answer of one question would lead to the next appropriate question, following the questioning flow chart presented in Figure (3).

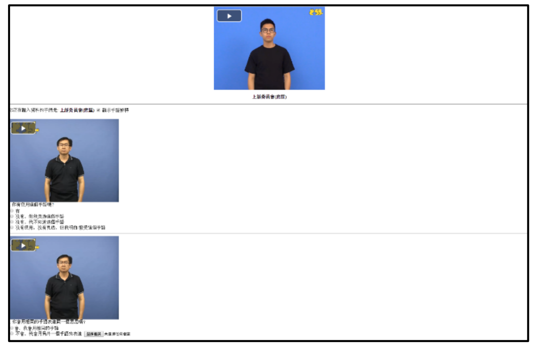


Figure 4: Screenshot of the Questionnaire Platform

4. Display of lexical variants in the Browser

After gathering all the data from the Questionnaire Platform, a Deaf researcher will help record the handshape features of the variants and the videos will be uploaded to HKSL Browser eventually. On the first page of the HKSL Browser, the purpose of the database (Figure (5)) and the explanation of lexical variations (Figure (6)) are offered in both written and signed language.

Figure 5: Introductory page of the HKSL Browser

Figure 6: Explanation of signing variation

Signs can be searched through written Chinese (i.e., no. of strokes involved), semantic categories (e.g., legal related concepts, transportation, education), and handshapes. The variants of each lexical target are represented by different colors, and they are ordered from left to right according to the degree of acceptability. Signs that are used and seen by more Deaf organizations are placed earlier. Figure (7) shows that there are five variants for the target concept “mini-bus”, represented by different colors.

Figure 7: A sample of the lexical variants

For each variant, viewers can check the result of the questionnaires by clicking the link below the video. Another link is also provided for people to offer their comments to the signs. Figure (8) shows the questionnaire result page.

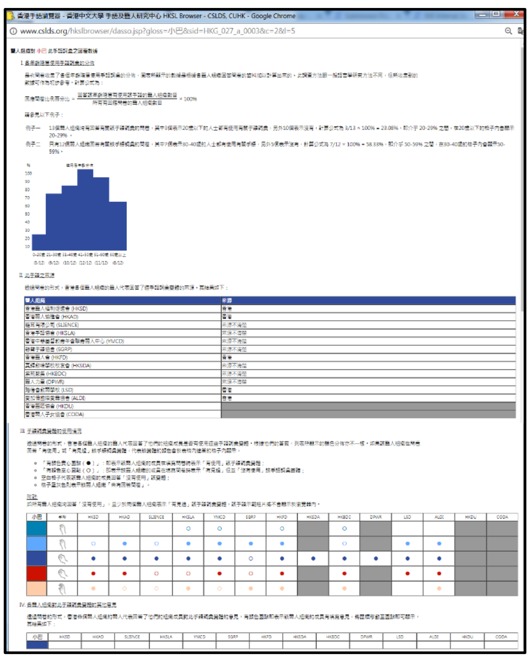


Figure 8: A sample of the questionnaire

5. Conclusive remarks

To sum up, the data collection and evaluation procedure of the HKSL Brower are designed to involve the Deaf signing community as much as possible and opportunities are provided for Deaf signers to contribute their sign language knowledge. Essentially, the HKSL Browser is the result of the concerted efforts of Deaf individuals as well as Deaf organizations in the community. It provides an open but monitored platform for Deaf signers to share and compare their lexical knowledge, which gradually helps to cultivate a sense of mutual respect and understanding among stakeholders with regard to lexical variations in HKSL. At present, there are 1,524 lexical targets with a total of 4,544 lexical variants. Since its first release in 2013, the Browser has been visited by 43,330 individuals (i.e., individual IP addresses). The total number of visits stands at 71,054, and the total number of sign viewing is 320,607. Every year CSLDS applies for government funding to expand the HKSL Browser. The new entries cover both daily signs and signs for specific purposes. For example, a specific set of signs related to the Olympic Games were added in 2016 to provide references for the TV interpreters for the sport events. In recent years, an increasing number of signs related to legal, social and political issues are added in response to the government’s commitment in increasing sign interpretation in public domains such as news and the Legislative Council meetings. Planning for other specialized areas is also underway for the sake of developing interpretation training programmes as well.

7. Acknowledgement

We would like to express our heartfelt thanks to the Hong Kong Special Administration Region Government for their financial support to the development of the Browser since 2012 (Implement Public Education Activities on Rehabilitation, The Labour and Welfare Bureau, The Hong Kong SAR Government). Our thanks also go to the Deaf signing informants for their participation in the data collection phases and all the organizations that have taken part in the evaluation processes. These organizations are (listed in no particular order): The Hong Kong Society for the Deaf, The Hong Kong Association of the Deaf, SILENCE, The Hong Kong Sign Language Association, The Chinese YMCA of Hong Kong Y's Men's Centre for the Deaf, The Silent Group, The Hong Kong Federation Of The Deaf, The Hong Kong School for the Deaf Alumni Association, The Hong Kong Bauhinias Deaf Club, The Hong Kong Deaf Empowerment, Lutheran School For The Deaf, The Agape Love Deaf International, The Hong Kong Deaf and Dumb Association, and The CODA Hong Kong.

8. References

Schembri, Adam, Jordon Fenlon, Ramas Rentelis, Sally Reynold & Kearsy Cormier. 2013. Building the British Sign Language Corpus. *Language Documentation and Conservation*, 7, 136-154.

Sze, Felix, Connie Lo, Lisa Lo & Kenny Chu. 2012. Lexical variations and diachronic change in Hong Kong Sign Language: preliminary observations. Paper presented in New Ways of Analyzing Variation in Asia Pacific 2, 1-4 August 2012, NINJAL, Tokyo.

Sze, Felix, Connie Lo, Lisa Lo & Kenny Chu. 2013. Historical development of Hong Kong Sign Language. *Sign Language Studies*, 13(2), 155-185.

Depicting Signs and Different Text Genres: Preliminary Observations in the Corpus of Finnish Sign Language

Ritva Takkinen, Jarkko Keränen, Juhana Salonen
University of Jyväskylä, Department of Language and Communication Studies, Sign Language Centre
P.O. Box 35, FI-40014 University of Jyväskylä, Finland
E-mail: ritva.takkinen@jyu.fi, jarkko.keranen@jyu.fi, juhana.salonen@jyu.fi

Abstract
In this article we first discuss the different kinds of signs occurring in sign languages and then concentrate on depicting signs, especially on their classification in Finnish Sign Language. Then we briefly describe the corpora of Finland's sign languages (CFINSL). The actual study concerns the occurrences of depicting signs in CFINSL in different text genres, introductions, narratives and free discussions. Depicting signs occurred most frequently in narratives, second most frequently in discussions and least frequently in introductions. The most frequent depicting signs in all genres were those that depicted the whole entity moving or being located. The second most frequent were those signs that expressed the handling of entities. The least frequent depicting signs were those with size- and shape-tracing handshapes. The proportion of depicting signs of all the signs in each genre was 17.9% in the narratives, 2.9% in the discussions and 2.2% in the introductions. In order to deepen the analysis, depicting signs will have to be investigated from the perspective of movement types and the use of one or two hands.

Keywords: Depicting sign, handshape, text genre, corpus, Finnish sign language

1. Introduction

1.1. Sign Types

Sign languages include several kinds of vocabulary. Not all signs are lexicalised (e.g. Brennan, 1992; Wallin, 1994; Brentari & Padden, 2001; Liddell, 2003; Johnston and Schembri, 2007, Takkinen, 2008). The visual-gestural modality that sign languages use to mediate linguistic information enables them to use space to gradually (analogously) convey visual content. The handshapes and movements in some signs can be motivated by inherent visual features and the movement or location of the entities in question.

According to Cogill-Koez, (2000), iconic signs or expressions of this kind are not linguistic but they are template visual representations that are on the continuum of analogous and schematic visual representations. Tolar et al. (2008) proposed that iconic signs can be pantomimic, perceptual, or both. Pantomimic signs primarily depict action-based features like KEY (open with a key). Perceptual signs depict static features like GLASSES. A sign like CAMERA can include both features, the action of taking a photograph and the size and shape of a camera. DeMatteo already in 1977 argued that morphemic analysis, which e.g. Supalla (1982) defended, is not suitable for these kinds of non-lexicalised signs. Liddell (2003) considers handshapes and some movement types lexical units, and other parameters (some movement types, articulation place, orientation and non-manual elements) gestural (analogous). Thus, in these signs lexical structure and depiction together create the meanings (also Emmorey & Herzig, 2003; Erlenkamp, 2009).

Johnston and Schembri (2010) divided the lexicon of Auslan into content signs and function signs, on the one hand, and to fully-lexical and partly-lexical signs, on the other. Fully-lexical signs constitute a listable lexicon in a sign language, and they may be either content signs or function signs. Fully-lexical signs include fully specified signs and partly specified (see also Johnson & Liddell, 1986; Johnston & Schembri, 2007). According to Johnston and Schembri (2010), the distinction between the two types of signs is more gradient than categorical.

1.2. The Corpora of Finland's Sign Languages (CFINSL)

In Finland there are two national sign languages, Finnish Sign Language (FinSL), with about 4000-5000 deaf users, and Finland-Swedish Sign Language (FinSSL), which has less than 100 deaf users. The CFINSL project aims to gather language data and create a machine-readable corpus of both sign languages. By the end of 2017, 92 FinSL users and 12 FinSSL users had been filmed in dialogue settings in a professional studio. The raw data were edited to video clips according to tasks and participants. The annotation process is going on with the ELAN program (Crasborn & Sloetjes, 2008) and annotation conventions have been (and will be) refined in the course of work.

In the CFINSL project, lexicalised signs are annotated on the ID tier without any additional code, and depicting, gestural, numeral, and fingerspelled signs with the following codes:

Lexicalised signs	HORSE
Depicting signs (_ds)	_dswe/_dshd/_dsss
Gestural signs (_g)	PALM-FORWARD_g
Numeral signs (_num)	TWO-WEEKS-AGO_num
Fingerspelling (_fs)	h-a-r-r-y_fs

Depicting signs are annotated following the classification of depicting verbs put forward by Takkinen (2008). The handshapes in these verbs are divided into three classes: 1) handshapes representing the whole entity, 2) handshapes representing handling an entity, and 3) handshapes depicting the size and shape of an entity (cf. Schembri, 1996; Cormier et al. 2012). Some researchers, e.g. Liddell & Johnson (1987) and Engberg-Pedersen (1993), have proposed a more detailed classification of handshapes depicting surfaces and extents of entities but Takkinen (2008) has combined them into one class of size- and shape-tracing handshapes. The movements are also divided into three categories according the classification proposed by Liddell & Johnson (1987): 1) a process movement represents the movement of an entity or the movement of an agent who is moving or touching an entity, 2) a contact movement indicates that an entity is located at a particular place, and 3) a tracing movement represents the surface or extent of an entity. The contact movement is a fixed type consisting of a short movement ending in a hold (MH type, see Liddell & Johnson, 1989; Liddell, 2003), and it occurs only with a whole entity handshape. Additionally, the tracing movement occurs only with a tracing handshape. Table 1 shows how handshape types and movement types can be combined.

Handshape / **Movement**	**Whole entity (we)**	**Handling (hd)**	**Size- and shape-tracing (ss)**
Process	A car ran on a hilly road. The sun is rising.	I picked up a leaf from the ground. The decorator smoothed the wallpaper.	
Contact	There is a car parked. There are apples on the tree.		
Tracing	There are cars in lines in the car park		The lake is still. The door has broad frames.

Table 1. Combinations of handshape classes and movement classes.

Figures 1a-1e illustrate the examples presented in Table 1. There is one example of the different combinations of a handshape class and a movement class.

Figure 1a. A car ran on a hilly road. (Combination of a whole entity handshape and a process movement: an entity is moving.)

Figure 1b. There is a car parked. (Combination of a whole entity handshape and a contact movement: an entity is located.)

Figure 1c. There are cars in lines in the car park. (Combination of a whole entity handshape and a tracing movement: several entities beside each other are seen like a surface.)

Figure 1d. I picked up a leaf from the ground. (Combination of a handling handshape and a process movement: an agent is moving an entity.)

Figure 1e. The door has broad frames. (Combination of a size- and shape-tracing handshape and a tracing

1.3. Corpus-based Research on Depicting Signs

Until now there has been little corpus-based research on depicting signs – the prevalence of different types of depicting signs or functions – in sign languages. One notable study, however, has been by Ferrara (2012) who, in her doctoral thesis, examined depicting signs in naturalistic Auslan conversation and narratives, concentrating on their function within clause-like units, their sign-level characteristics and their interaction with constructed action.

Ferrara (2012) compared conversation and narrative data and found that the narratives contained twice as many depicting signs as did the conversation data. Ferrara pointed out that the elicitation material may have influenced the results. In the conversation data there was no visual or other material but in the narrative data the material was visual, not linguistic. Purely visual material may motivate signers to use more depicting signs in their signing. In addition, in Ferrara's study 82.2% of the depicting signs were produced with two hands, either two-handed symmetrical or asymmetrical. Most often the non-dominant hand was in the background participating in the depicted action being carried out by the dominant hand. (Ferrara 2012.)

In this analysis, our research question was to what extent depicting signs occur in different text genres: introductions, narratives and discussions.

2. Data and Methods

The data of this analysis consist of the signing of 22 informants who are all early signers, i.e. they are the deaf or hearing children of deaf signing parents or they are deaf children who have acquired sign language as their first language in early childhood (or some older deaf persons at school age). The informants are between 18 and 84 years of age. Table 2 shows the age groups and the number of informants in each group. Twelve of the informants are men and ten are women.

Age group	Number of informants
18-29	4
30-39	5
40-54	1
55-69	7
70 -	5
Total	**22**

Table 2. The number of informants in different age groups.

Most of the informants are from the central part of Finland, one is from eastern Finland and one from southern Finland.

The data were filmed in a professional studio setting with six cameras; one camera recording a general view, two recording the complete picture, two a closer picture of each interlocutor, and one recording the interlocutors from above. Two informants were interacting with each other in each session, led by a native signer.

The sign language data were gathered by giving the informants seven different language tasks: 1) an introduction, 2) a discussion of work or hobbies, 3) narrating about cartoon strips (Ferd'nand), 4) narrating about a video, 5) narrating a story from a picture book (The Snowman, and Frog, where are you?), 6) discussing a topic related to the deaf world, and 7) free discussion (e.g. on travelling, TV programmes, sports). Tasks 1–2 and 6–7 are discussions, and tasks 3–5 are narrative monologues, but the other interlocutor was able to put comments or questions during the narration.

For this research tasks 1) introduction, 5) narrating a story from a picture book, and 6) discussing a topic related to the deaf world were analysed. The introduction data includes talking about the participants' name signs, their childhood, where each of the participants was born and where they went to school, as well as their family background. Before narrating from the picture books the participants had time to go through the books and gather their thoughts. When narrating they no longer looked at the book. Discussing a topic related to the deaf world was a free discussion without any elicitation material. The total length of the data is seven hours: task 1 is 3 hours, task 5 is 1.5 hours, and task 6 is 2.5 hours. The total number of tokens is 43,532.

In our corpus project we have decided to code depicting signs on the ID tier according to the handshape classes presented above. Depicting signs are annotated with the code _*ds* (_kv in Finnish). The different classes of handshapes are separated with the codes *we* (whole entity), *hd* (handling), and *ss* (size and shape) (See Figure 1)[1]. In order to explore depicting signs according to the movement types, an extra annotation tier will have to be added; this is easy to create later because, the depicting signs are already identified on the ID tier. In the Auslan corpus, for example, depicting signs are grouped into four sub-types: signs depicting movement, location, handling and size and shape (Johnston 2016; Ferrara 2012).

Figure 1. A snapshot of the ELAN screen.

In the annotation process every sign is estimated as to whether it has a lexicalised form or not and what kind of function it serves in the signing text. If it cannot be glossed with a fixed gloss and it serves a depicting (predicative) function in the text, it is annotated as a depicting sign.

To support annotation, a web-based lexical database, Signbank, originally created for the Auslan corpus[2],[3], was created for the CFINSL (Salonen et al. 2016). The glosses in Signbank are exported to ELAN via ECV (externally controlled vocabulary). This helps to keep the annotation conventions consistent and makes the annotation easier and quicker.

3. Results

In our data, depicting signs occurred most frequently in narratives (1413), second most frequently in discussions (500) and least frequently in introductions (413). Table 3 shows the frequencies in detail.

DS types	Introduction	Narrative	Discussion	Total
dswe	146	658	253	**1057**
dshd	135	479	128	**742**
dsss	132	278	119	**529**
Total	**413**	**1413**	**500**	**2328**

Table 3. The number of depicting signs in different genres.

If we look at the frequencies of different types of depicting signs, those that depicted the whole entity moving or being located were the most frequent in all genres (*dswe*). The second most frequent were those signs that expressed the handling of entities (*dshd*). The least frequent depicting signs were those with size- and shape-tracing handshapes (*dsss*). The difference between the two groups *dshd* and *dsss* was largest in the narratives. On the other hand, in the introduction and the discussion data they occurred almost equally often.

Figure 3 shows the prevalence of the different classes of handshapes in a more visual form.

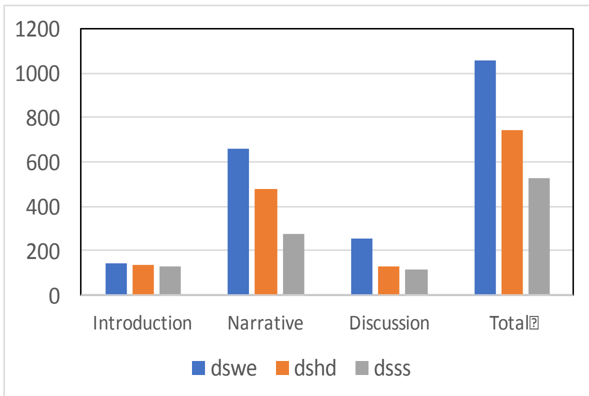


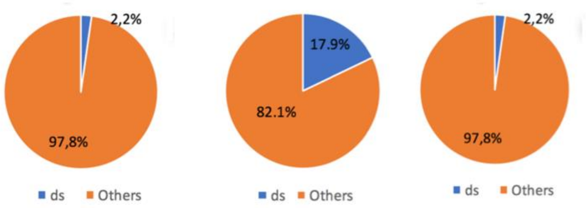
Figure 3. Occurrences of different handshape types in depicting signs in different genres.

Figure 4 displays the percentages of depicting signs in different genres. Depicting signs occurred most often in the narratives (17.9%) in relation of all signs in this genre. In the discussion only 2.9% and in the introductions only

[1] Cf. Johnston (2016) Auslan Corpus Annotation Guidelines: type-like information precedes token-like infor- mation.

[2] Auslan Signbank http://www.auslan.org.au
[3] CFINSL Signbank has been developed on the basis of the NGT Signbank http://signbank.science.ru.nl

2.2% of all signs were depicting signs. In the corpus-based study of de Breuzeville et al. (2009) the proportion of depicting signs in narratives was 9%. In BSL conversation data the prevalence of classifier signs was 2.3% (Fenlon et al. 2014) and in Auslan casual free conversation it was 7.3 % (Johnston 2012).



a b c

Figure 4. The percentages of depicting signs in a) the introductions, b) the narratives, c) the discussions.

An example from the FinSL corpus of depicting signs including different handshape classes is presented in Figure 5. This clip is from the Snowman story. The signer uses both hands while depicting how the Snowman (*RH:dswe*) and the boy (*LH:dswe*) are about to leave the ground and then they fly upwards (*dswe*). The Snowman (*RH:dswe*) is holding (*LH:dshd*) the boy's hand. While flying (*dswe*) they look down and the Snowman sees the surface (*RH:dsss*) of the earth, and he holds (*LH:dshd*) the boy's hand until they come down again (*dswe*).

RH: dswe LH: dswe 'the beings are preparing to fly'	RH: dswe LH: dswe 'the beings are flying'	RH: dswe LH: dshd 'the being is flying and holding the other one by hand'

RH: dswe LH: dswe 'the beings are flying'	RH: dsss LH: dswe 'being sees the ground and holds the other one in hand'	RH: dswe LH: dswe 'the beings are landing'

Figure **5**. Example of depicting signs including different handshape types.

4. Discussion

The results show that the prevalence of depicting signs is highest in narratives. That was shown also by Ferrara (2012) and in other earlier studies (e.g. Morford & Macfarlane, 2003; Johnston, 2012), and it is similar to everyday experience in sign language use. The introduction and discussion data showed a low number of depicting signs even though the duration of their data was twice as long as that of the narratives. On the other hand, the signing speed may have been quicker and the production smoother in narratives compared to introductions and discussions, which were interrupted by the interlocutor's comments and questions.

The visual elicitation material – as Ferrara (2012) noted – may have affected the notably higher number of depicting signs in narratives. The topic of the discussion may also affect how much and what kinds of depicting signs appear there (e.g. Keränen, 2017). It is an interesting and still open question whether the quality of the narrative in terms of whether it is about a private experience or talking about other people affects the frequencies of different types of depicting signs.

The classification of depicting signs – and even the terms used for that kind of sign – varies from one researcher to another, which affects the annotation conventions. Additionally, whatever kind of classification is created, the decision about what is annotated as a depicting sign is not always easy. All this makes it more difficult to make comparisons between different datasets or corpora. The comparison of frequencies can only be approximate. Nevertheless, corpora will make it much more efficient to carry out cross-linguistic studies than it has been with separate small datasets.

In order to study depicting signs in FinSL (or in FinSSL) in more detail we need to create additional tiers, e.g. from the perspective of the movement of the depicting signs, as well as to analyse the use of one or two hands and their functions. It would be interesting to study the contexts and iconicity (pantomimic, perceptual or both) of depicting signs in different genres, i.e. what is behind the frequencies. A more detailed analysis of depicting signs as well as of other partly-lexical signs is important in the description of sign languages. The more we know about the structural potential and function of these signs, the better we can contribute to knowledge about human languages and the better we can teach sign language as a mother tongue to early signers and as a foreign language to foreign language learners, e.g. to the hearing parents of deaf children.

5. Acknowledgements

The authors wish to thank the informants, the S. and A. Bovallius Foundation, The Swedish Cultural Foundation in Finland and the Ministry of Education and Culture for financial support for the CFINSL project as well as the reviewers for the comments on the manuscript.

6. Bibliographical References

Brennan, M. (1992). The visual world of BSL. In D. Brien (Ed.) Dictionary of British sign language / English. London, Boston: Faber and Faber, pp. 1–133.

Brentari, D. & Padden, C. (2001). Native and Foreign Vocabulary in American Sign Language. In D. Brentari (ed.), *Foreign Vocabulary in Sign Languages*. Mahwah, NJ: Lawrence Erlbaum, pp. 87–119.

de Beuzeville, L., Johnston, T. & Schembri, A. (2009). The use of space with indicating verbs in Auslan. A corpus-based investigation. Sign Language & Linguistics 12:1, 53–82. doi 10.1075/sl&l.12.1.03deb

Cogill-Koez, D. (2000). A model of signed language 'classifier predicates' as templated visual representation. Sign language & linguistics 3, pp. 209–236.

Crasborn, O. & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages. Paris: ELRA, pp. 39–43.

DeMatteo, A. (1977). Visual imagery and visual analogues. In L Friedman (Ed.), On the other hand: Resent perspectives on American sign language. New York: Academic Press, pp. 109–136.

Emmorey, K. & Herzig, M. (2003). Categorical versus gradient properties of classifier constructions in ASL. In K. Emmorey (Ed.), Perspectives on Classifier Constructions. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., pp. 199–220.

Erlenkamp, S. (2009) "Gesture verbs". Cognitive-visual mechanism of "classifier verbs" in Norwegian Sign Language. *CogniTextes*, Volume 3/2009. DOI : 10.4000/cognitextes.250

Ferrara, L. (2012). The grammar of depiction: Exploring gesture and language in Australian Sign Language (Auslan). Doctoral thesis. Department of Linguistics, Macquarie University, Sydney, Australia.

Johnston, T. & Schembri, A. (2007). Australian Sign Language (Auslan): An introduction to sign language linguistics. Cambridge: Cambridge University Press.

Johnston, T. & Schembri, A. (2010). Variation, lexicalization and grammaticalization in signed languages. Langage et société, 131, 19–35. DOI: 10.3917/ls.131.0019

Johnston, T. (2012). Lexical frequency in signed languages. The Journal of Deaf Studies and Deaf Education, 17(2), 163–193.

Johnston, T. (2016). Auslan Corpus Annotation Guidelines. http://www.academia.edu/29690332/Auslan_Corpus_Annotation_Guidelines_November_2016_revision_

Keränen, J. (2017). Käsiotteen ja viittomakielisen käsimuodon välinen ikonisuus kognitiivis-semioottisesta näkökulmasta. (Iconicity of hand grasp and handshape of signed language from the cognitive-semiotic perspective.) Master's thesis. Jyväskylän yliopisto, Kieli- ja viestintätieteiden laitos. https://jyx.jyu.fi/dspace/bitstream/handle/123456789/52903/URN%3aNBN%3afi%3ajyu-201701301315.pdf?sequence=1

Liddell, S. & Johnson, R. (1987). An analysis of spatial-locative predicates in American sign language. Paper presented at the fourth international conference on sign language research, 15–19 July, Lappeenranta, Finland.

Liddell, S. (2003). Grammar, gesture, and meaning in American sign language. Cambridge: Cambridge University Press.

Morford, J. P. & Macfarlane, J. (2003). Frequency Characteristics of American Sign Language. Sign Language Studies, 3 (2), 213–225.

Salonen, J., Takkinen, R., Puupponen, A., Nieminen, H. & Pippuri, O. (2016). Creating Corpora of Finland??s Sign Languages. In E. Efthimiou, F. Stavroula-Evita, T. Hanke, J. Hochgesang, J. Kristoffersen & J. Mesch (Eds.), Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining [organized as a part of LREC'16 at Portoroz, Slovenia, May 28, 2016]. Paris: European Language Resources Association (ELRA), pp. 179–184.

Schembri, A. (1996). The structure and formation of sign in Auslan (Australian sign language). Renwick College Monograph No. 2.

Supalla, T. (1982). Structure and acquisition of verbs of motion and location in American Sign Language. P.h.D. thesis, University of. California.

Takkinen, R. (2008). Kuvailevat verbit suomalaisessa viittomakielessä. (Depicting verbs in Finnish Sign Language.) Puhe ja kieli 28:1, 17–40.

Tolar, T. D., Lederberg, A. R., Gokhale, S. & Tomasello, M. (2008). The Development of the Ability to Recognize the Meaning of Iconic Signs. Journal of Deaf Studies and Deaf Education, 13, 225–240.

Wallin, L. (1994). Polysyntetiska tecken i svenska teckenspråket (Polysynthetic signs in the Swedish Sign Language). Doctoral dissertation. Institutionen för lingvistik, Avdelningen för teckenspråket, Stockholms universitet.

7. Language Resource References

ELAN. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. URL: http://tla.mpi.nl/tools/tla-tools/elan.

Improving Lemmatisation Consistency without a Phonological Description The Danish Sign Language Corpus and Dictionary Project

Thomas Troelsgård, Jette Hedegaard Kristoffersen
Centre for Sign Language, University College Copenhagen
{ttro, jehk}@ucc.dk

Abstract

The Danish Sign Language Corpus and Dictionary project at Centre for Sign Language, UCC has a dual aim: to build of Danish Sign Language Corpus, and to use this corpus to expand and improve The Danish Sign Language Dictionary. Our goal is a one-to-one correspondence between sign lemmas in corpus and dictionary, but due to limited resources, we cannot include an accurate phonological description of each sign form. In order to secure a consistent lemmatisation in the corpus as well as across the two resources, we thus rely exclusively on sign videos and Danish equivalents. In this paper, we will describe how we use the lemmas of the Danish Sign Language Dictionary, and additional signs found in connection with the dictionary work, as the initial lexical database of the corpus tool. For new signs found in corpus, the actual corpus tokens will serve as preliminary video representations. To facilitate the sign search when lemmatising corpus tokens, we assign several Danish equivalents to each sign, including all equivalents in the dictionary data. Furthermore, we include synonyms found through linking these equivalents to the Danish wordnet (DanNet), although equivalents added in this way cannot be regarded as valid senses of the sign.

Keywords: corpus linguistics, annotation, sign language, language documentation, Danish Sign Language (DTS)

1. Introduction

The Danish Sign Language Corpus and Dictionary project is carried out at the Centre for Sign Language at UCC - by the same project group that developed the Danish Sign Language Dictionary (Ordbog over Dansk Tegnsprog; cf. Kristoffersen and Troelsgård, 2012). In 2015, we began working on a corpus of Danish Sign Language (DTS), the first of its kind. The current project phase has a dual goal: to build a corpus of DTS, and to expand and improve The DTS Dictionary based on this new corpus. For building our corpus we use the iLex system (cf. Hanke and Storz, 2008), a database tool that is developed at the University of Hamburg.

In order to secure consistency across corpus and dictionary, we aim at a one-to-one correspondence between the dictionary lemmas and the corpus lexicon – the set of types used for lemmatising corpus tokens. Unique identifiers of sign types are essential to machine readable text that can serve as the source for linguistic analysis of the sign languages (cf. Johnston, 2010). The lack of a written standard for Sign Languages commonly used by native signers complicates the identification of the lemmas in the annotation process (cf. Zwitserlood et al., 2013). To achieve an unambiguous lemmatisation, some corpus projects, e.g. the German Sign Language Corpus (DGS-Corpus), include a detailed formal description of the sign form, e.g. in HamNoSys (The Hamburg Sign Language Notation System, cf. Hanke, 2004). Other projects represent a sign solely through a gloss – typically a word from the surrounding spoken language, chosen as a mnemonic because it captures (one of) the core meaning(s) of the sign. For the DTS Corpus project we chose to use only glosses because of limited resources. In this paper, we will describe how we try to achieve a high degree of consistency in the corpus annotation and across corpus and dictionary, without having a searchable description of the sign form.

2. Building the vocabulary

In the iLex system, the lemmatisation task is performed by linking every token to a matching type in the lexical database. Obviously, this linking is completed faster, easier and more reliable if the initial sign vocabulary is large and well described, ideally having both a video, a searchable formal form description (e.g. HamNoSys or Stokoe), and one or more spoken language equivalents. Because of limited funding, we decided to leave out the formal description, and go with only videos and Danish equivalents (and/or a prose description of function or use).

2.1 Initial vocabulary

For building our sign vocabulary, we first included the approximately 2.200 lemmas of the DTS Dictionary. As the signs were already analysed regarding form and meaning,, we decided to re-use the definitions of homophony and phonological variation that we use for the dictionary (Kristoffersen and Troelsgård, 2012), and hence (ideally) end up with a one-to-one relation between sign units in the dictionary and in the corpus project.

As a tool for lemma selection for the dictionary project, we built a database containing the signs from a number of older dictionaries and sign lists. We then began analysing video recordings of DTS provided by our group of consultants, adding new signs to the database as we encountered them in the videos. The database was then used as source for the selection of lemmas for the DTS Dictionary. During the following lexicographic work on the dictionary, new signs were continuously added to the database. While building the sign type vocabulary for the corpus, we included all signs from the database that were not already dictionary lemmas. In connection with adding signs to the database, we also added the known phonological variants of each sign according to the variant definition of the dictionary: signs with the same semantic content and variation in only one of the major phonological parameters: handshape, orientation, movement, place of articulation, are regarded as phonological variants of one sign (cf. Troelsgård and

Kristoffersen, 2008). Finally, as a preparation for the corpus project, we made studio recordings of all signs and their phonological variants in the database that were not already dictionary lemmas.

Consequently the initial sign vocabulary in the corpus system consisted of about 7.000 signs (and about 1.000 additional sign variants), all accompanied by a video recording (either from the dictionary or added in connection with the preparation of the corpus project).

2.2 Adding new signs

As soon as we started annotating corpus videos, obviously the need occurred of being able to add new signs to the vocabulary as we encounter them. These signs are lemmatised using temporary "dummy signs", which are regularly checked, and – if they are found actually to be missing in the vocabulary – added to the database, with the actual corpus tokens serving as video evidence. All signs found in the corpus are regarded as future lemma candidates for the dictionary. If a sign is later selected as a dictionary lemma, we will compile a new entry based on an analysis of the corpus tokens, and we will make studio recordings of the sign and its variants.

3. Adding equivalents

As we decided not to include a formal phonological description, it is essential to provide one or more Danish equivalents to each sign. As the 2.200 dictionary signs already were semantically analysed, and described as having one or more sense (each with one or more Danish equivalents, and/or a prose description of function or use), we decided to exploit the possibility in the iLex system of structuring the sign type vocabulary as a hierarchy, and thus we clustered the equivalents according to the word-senses defined in the dictionary. As a result, we work with a three level hierarchy, which we will illustrate through the sign FRUIT, a sign described as having two word-senses, and two phonological variants. The variants differ in handshapes – the movement is in both cases a twist of the wrist, see Figure 1.

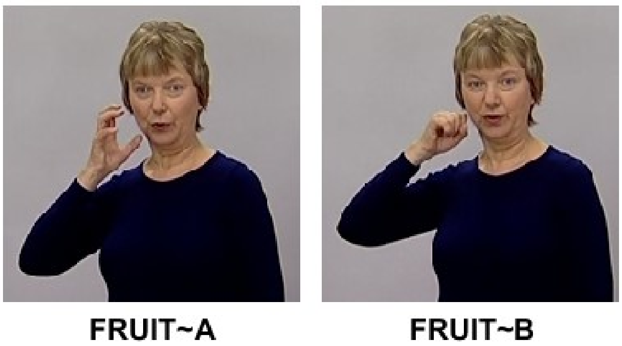


Figure 1: The two phonological variants of the DTS Sign FRUIT.

For the type hierarchy this gives one type at the sign level, two types at the variant level (form), and four types at the meaning level (combination of form and sense), as shown in Figure 2. A more detailed description of the way we use the iLex type hierarchy can be found in Langer et al. (2016).

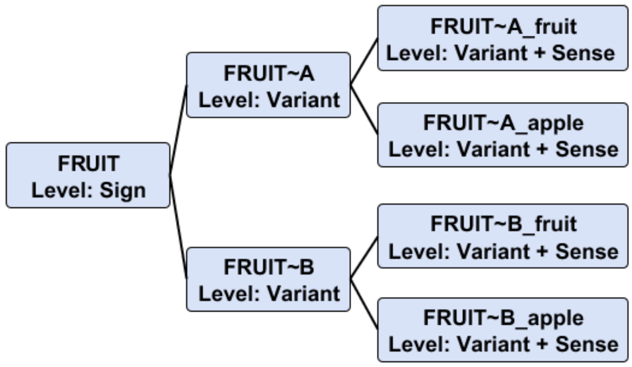


Figure 2: The three-level sign type hierarchy used in the iLex system for the DTS Corpus project.

At the meaning level, we add the first equivalent of the corresponding dictionary word-sense to the gloss as a disambiguator. Furthermore, we use iLex' module for linking types with concepts to assign all Danish equivalents from the DTS Dictionary to the type, thereby making it possible to find the sign through these equivalents. As an example of this linking, we use the sign WOMAN. The DTS Dictionary entry of WOMAN is shown in Figure 3.

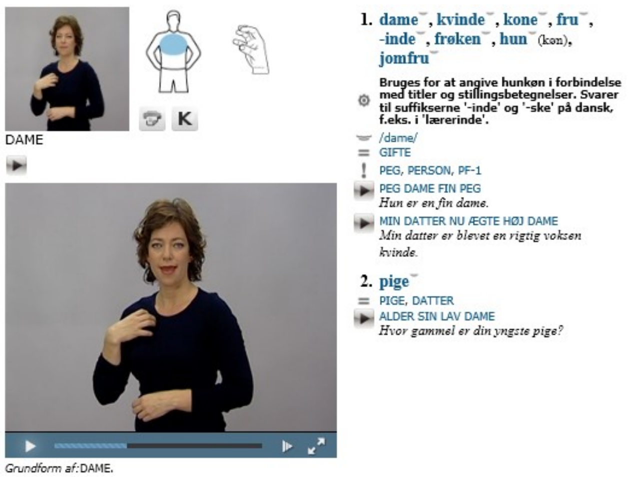


Figure 1: DTS Dictionary entry of WOMAN.

Table 1 shows the meaning level types of the two senses of WOMAN, and the linked (and searchable) equivalents taken from the DTS Dictionary.

Type at meaning level	Linked equivalents
Sense 1: WOMAN_woman	dame (woman) kvinde (wife) kone (wife) fru (madam) -inde (-ess) frøken (miss) hun (female) jomfru (virgin)
Sense 2: WOMAN_girl	pige (girl)

Table 1: The two meaning level types of WOMAN and their linked equivalents from the DTS Dictionary

4. Linking to DanNet

We wanted to add even more relevant Danish equivalents to each sign sense, thereby increasing the possibility of a match when searching signs through words. For this purpose, we chose to use the Danish wordnet, DanNet (DanNet; cf. Pedersen et al., 2009; Trap-Jensen, 2010). A wordnet is a semantic network that clusters closely related word-senses (synonyms and near-synonyms) into so-called synsets, and links these together according to semantic relations such as hyponymy, hypernymy, metonymy, entailment etc. We matched our dictionary equivalents against the DanNet words, and performed a semiautomatic linking between dictionary senses and relevant DanNet synsets. Using these links, we then were able to add equivalents to each word-sense, by including all synonyms of its linked DanNet synsets. Thus, if we consider the sign WOMAN, it is described in the DTS Dictionary as having two senses: ‘woman, wife’ and ‘girl’. The first sense has a number of equivalents in the dictionary data, including *dame* (‘lady’), *kvinde* (‘woman’), *kone* (‘wife’), *fru* (‘madam'). If we match e.g. *kone* (‘wife’), to DanNet, we get five additional equivalents from the synset of *kone*: *ægtehustru*, *ægteviv*, *frue*, *hustru* and *viv* (all meaning ‘wife’). When choosing equivalents for the DTS dictionary, we balanced word frequency against the total number of equivalents, and because of the large number of relevant equivalents for the sense ‘woman, wife’, none of the five words found through DanNet were chosen as equivalents for the entry WOMAN. Nevertheless, as shown in Table 2, the two most frequent words added through the DanNet matching: *hustru* and *frue* are fairly frequent, and are likely to be used as search words during the lemmatisation of corpus tokens.

Danish equivalents	Frequency	DTS Dictionary
kvinde	3090	present
kone	2573	present
dame	942	present
hustru	579	absent
frue	184	absent
viv	12	absent
ægteviv	2	absent
ægtehustru	0	absent

Table 2: Danish words meaning 'wife', with word frequency count from the Korpus 90 Project[1]

Obviously, the equivalents added in this way cannot be regarded as valid senses of the sign – they are included solely for the purpose of increasing the opportunity of finding a sign though a word-based search. The possible sign-senses – and their appropriate equivalents – can only be deduced through analysis of the actual corpus tokens of each sign.

5. Word-based type search

In the absence of a formal sign description, word-based search is the primary means of identifying the correct sign type while annotating the texts of the DTS Corpus. Through a text search, hopefully the matching sign – checked by watching the connected video evidence – is found (preferably in a matching word-sense), and used for the lemmatisation.

A disadvantage of this approach is that it is impossible to foresee all possible search strings; hence, sometimes searches for signs that are actually in the system do not give any result. In these cases, we lemmatise using special dummy types. Later on, we examine these dummies, in order to decide whether they are instances of existing signs, or of new signs, not yet entered as types in the system.

Sometimes a search results in finding the appropriate sign, but not finding an adequate type at the meaning level. In these cases, we go up one level in the type hierarchy, lemmatising to a type at the variant level, e.g. using the type FRUIT~B, as shown in Figure 2, and indicating that the sign form is right, but the actual sense is neither ‘fruit’ nor ‘apple’.

6. Concluding remarks

For the DTS Corpus project, we do not have searchable formal sign descriptions at hand. Instead, we have chosen an approach where we add many spoken language equivalents to each sign, in order to increase the probability of finding the right sign when lemmatising corpus tokens. Furthermore, we work with a lexical sign base, where every record is represented by a video recording. This secures a correct choice of sign type. Especially when dealing with

[1] Korpus 90 was part of the work on Danish text corpora (cf. KorpusDK) carried out at Society for Danish Language and Literature (DSL, cf. www.dsl.dk). Recent word frequently lists from DSL can be downloaded at korpus.dsl.dk

phonological variants and sign synonyms, the video evidence secures a correct choice.

We believe that this approach is a feasible, second-best solution for sign language corpus projects without resources for performing a detailed phonological description of the sign vocabulary and tokens of their corpus. We also suppose that including the relation links of wordnets might increase the success rate of word searches, as might the inclusion of other spoken language resources, e.g. corpus tools for finding related words.

7. Acknowledgements

The database tool iLex has kindly been placed at our disposal by the developers at Hamburg University. We would also like to thank the DGS Corpus group at Hamburg University for ongoing support of our use of iLex and our corpus building.

The Danish Sign Language Corpus and Dictionary project is supported by funding from the "Diversity and Social Innovation" research fund of UCC (Professionshøjskolen University College Capital, Denmark), and from The Jascha Foundation (Denmark).

8. Bibliographical References

Hanke, T. (2004). HamNoSys - representing sign language data in language resources and language processing contexts. In O. Streiter, C. Vettori (Eds.): From SignWriting to Image Processing. Information techniques and their implications for teaching, documentation and communication. Proceedings of the Workshop on the Representation and Processing of Sign Languages. 4th International Conference on Language Resources and Evaluation, LREC 2004, Lisbon. Paris: ELRA, pp. 1-6.

Hanke, T. and Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. Construction and Exploitation of Sign Language Corpora. In Crasborn, O. et al. (Eds.) Construction and Exploitation of Sign Language Corpora. Proceedings of 3rd Workshop on the Representation and Processing of Sign Languages. ELRA, Paris, pp. 64-67.

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15(1):106–131

Kristoffersen, J. H. & Troelsgård, T. (2012). The electronic lexicographical treatment of sign languages: The Danish Sign Language dictionary. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography*. Oxford: Oxford University Press, pp. 293-318.

Langer, G., Troelsgård et al. (2016). Designing a Lexical Database for a Combined Use of Corpus Annotation and Dictionary Editing. In Efthimiou, E., Fotinea, E. et al. (Eds.) Corpus Mining: Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages. 10th International Conference on Language Resources and Evaluation, LREC 2016, Portorož. Paris: ELRA, pp. 143-152

Pedersen, B.S, Nimb S, Asmussen J. et al., (2009). DanNet – the challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. *Language Resources and Evaluation,* 43(3):269-299. Springer, Netherlands.

Zwitserlood, I, Kristoffersen, J and Troelsgård, T (2013). Issues in Sign Language Lexicography. In *Jackson, H. (ed.): The Bloomsbury Companion to Lexicography*, pp. 259–283. London: Bloomsbury Publishing.

Trap-Jensen, L. (2010). Access to multiple Lexical Resources at a Stroke: Integrating Dictionary, Corpus and Wordnet Data. In Granger, S. & Paquot, M. (Eds.), Lexicograhy in the 21st Century: New Challenges, New Applications. Proceedings of eLex 2009, Louvain-la-Neuve. pp 295-302. Cahiers du CENTAL, Louvain-la-Neuve: Presses Universitaires de Louvain.

Troelsgård, T. and Kristoffersen, J.H.(2008) An electronic dictionary of Danish Sign Language. In Müller de Quadros, R (Ed.) Sign Languages: spinning and unraveling the past, present and future. TISLR9, forty-five papers and three posters from the 9th Theoretical Issues in Sign Language Research Conference Florianopolis, Brazil, pp. 352-362. Editora Arara Azul.

9. Language Resource References

Corpus Resources & Documentation. Society for Danish Language and Literature, DSL. Accessed at: http://korpus.dsl.dk/ [20/02/2018]

DanNet [The Danish wordnet]. Accessed at: http://wordnet.dk [28/11/2017]

DGS-Corpus [German Sign Language Corpus]. Accessed at: http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/welcome.html [9/1/2018]

KorpusDk [The Danish Corpus] Accessed at: http://ordnet.dk/korpusdk [20/02/2018]

Ordbog over Dansk Tegnsprog [The Danish Sign Language Dictionary]. Accessed at http://www.tegnsprog.dk [9/1/2018]

Hand in Hand – Using Data from an Online Survey System to Support Lexicographic Work

Sabrina Wähl, Gabriele Langer, Anke Müller
Institute of German Sign Language and Communication of the Deaf, University of Hamburg
Gorch-Fock-Wall 7, 20354 Hamburg, Germany
E-mail: {sabrina.waehl, gabriele.langer, anke.mueller}@uni-hamburg.de

Abstract

In the DGS-Korpus project the lexicographic descriptions of signs are based on available data of the DGS-Korpus, a reference corpus of German Sign Language (DGS). As this corpus is limited in size, number of informants recorded and topics included, it is in some cases helpful to obtain additional information from the larger sign language community via an online voting system. This is done using the DGS-Feedback System, a tool especially designed for online surveys conducted using a sign language. With this tool further information on e.g. sign forms and meanings and their use and regional distribution has been elicited. Data from the DGS-Feedback is used in several ways during the lexicographic process of preparing dictionary entries to supplement data from the corpus. In the following the consideration of the DGS-Feedback data in relation to the corpus data in decision-making, analysis, and lexicographic description is explained and discussed by way of examples.

Keywords: corpus-based lexicography, online voting system, community sourcing, German Sign Language

1. Introduction and Background

New technologies have made it possible to build sign language corpora of considerable sizes. The DGS-Korpus project has now a corpus consisting of 560 hours of recorded signed communication of which approx. 465.000 tokens have been annotated (23.02.2018). Nevertheless, this corpus is limited in size, in number of informants recorded and to the topics that were included as elicitation stimuli (Hanke et al., 2010; Nishio et al., 2010) or that came up spontaneously during the recorded conversations.
Within the DGS-Korpus[1] project an online survey tool, the DGS-Feedback System, was developed to facilitate the use of a sign language throughout the survey for asking and answering questions and giving controlled comments (König et al., 2013; Langer et al., 2014). It was developed to address the DGS community, but could also be used for other sign languages (open source). Within the project this tool was first used for surveys to verify signs and their presumed meanings in previously published sign collections (Langer et al., 2014). Currently, the DGS-Feedback System is primarily used to supplement corpus data to be reviewed in the analysis stage when compiling corpus-based dictionary entries.
In the following we discuss how the data obtained through the DGS-Feedback are used and how they can help to complete the picture of a sign's use in addition to a corpus-based analysis.

2. Data from the DGS-Feedback

2.1 Sampling

The DGS-Feedback is open to all members of the DGS community who want to participate (Langer et al., 2014; Langer et al., 2016a). All participants fill out an initial questionnaire with information on their person and sign language use (metadata). This is needed for the analysis and interpretation of the results. Up to now, 279 persons (23.02.2018) have contributed to the DGS-Feedback. The sampling of the DGS-Feedback is subject to chance and therefore the group of language users participating is very heterogeneous including early and late learners, CODA, deaf, hearing, hard of hearing, and different age groups (Langer et al., 2016a). This is an important difference to the corpus, where the sampling of informants is balanced for gender, age group, and region. Also, all informants of the DGS-Korpus are native or near-native signers, as early learners were preferred over late learners. When using the data from the DGS-Feedback this heterogeneity of contributors has to be considered and weighted in the analysis.

2.2 Structure of the Survey

Different question types were developed to focus on different aspects of signs and sign use. In the first question type one sign form is presented to the user in combination with several meanings.[2] The second question type presents one concept and asks for different signs that are used for that meaning. For the purpose of this paper we will focus on the first question type. Questions of the first question type were the first to be released and are the first new participants are given to fill out before they can progress to the next level with the second question type. The goal of question type 1 is to check which meanings of a sign are used, known or unknown within the language community and to acquire more data on regional distribution.
In general a questionnaire (hereafter work package) consists of several question pages (hereafter questions). A question may include several question items. Within a question of the question type 1 first the respective sign is shown without mouthing and the participants are asked whether they know the presented sign form or not. If they know the form and chose *yes* further question items concerning the sign's meanings are presented. For each

[1] http://dgs-korpus.de, last access: 23.02.2018

[2] In the context of the DGS-Feedback we use the term *meanings* to refer to linguistic knowledge (on a sign) and with regard to corpus data to refer to the contextual meaning of an actual token. We use the term *sense* when it comes to the lexicographic analysis and description of such meanings, as it implies taking context patterns and actual use into account and describing them in a summarised way as a list of senses a sign can cover.

meaning the following stimuli are given: 1) a video clip of the single sign produced with a corresponding mouthing or mouth gesture, 2) written German equivalents, sometimes followed by a disambiguating written hint in brackets, and 3) in some cases a video clip with a signed context. In most cases the DGS context consists of a competence example of the sign. A DGS context is added in cases where the written information alone seemed insufficient or the German equivalents may not be well-known. A DGS context is also shown in cases where the distinction between closely related meanings has to be made particularly clear, and in cases where rather peripheral meanings are contrasted with presumed core meanings.

Figure 1: Stimulus and answer-buttons for one form-meaning combination in the DGS-Feedback

For each form-meaning combination the participants have the choice between three responses (see figure 1), which are: 1) *I use it* myself, 2) *I know it* from other signers, but do not use it myself or 3) *it is unknown to me*. In this paper these answers are referred to as *used*, *known* and *unknown*[3]. The answer *known* is the response option to select when participants are aware of an existing sign that they normally do not use themselves (passive vocabulary). At the end of each question concerning one sign form, the user is asked whether they miss a meaning they would like to bring to our attention. Answers can be given in writing or sign via a webcam. Once a work package is completed it can be submitted to the project. The results of returned work packages are imported into iLex[4] and can be analysed through queries and special list views. Up to date we released 42 work packages of type 1 of which 14 work packages with 71 different sign forms have more than 100 returns (23.02.2018).

3. Analysis Stage of Corpus-based Lexicographic Work

With a growing corpus and higher numbers of tokens per type available we have started with what Atkins and Rundell (2008:98-103) have called the *analysis* stage of dictionary making, that is, to analyse the available data of the sign in question and to document relevant facts about it. Central to this lexicographic work is the description of the sign's meanings and uses and grouping them into senses and sub-senses, a step sometimes called Word Sense Disambiguation (WSD) (cf. Atkins & Rundell, 2008:269). Basically, this is done by reviewing a substantial number of tokens in context, determining their contextual meaning and conditions of use, grouping these uses and describing them as senses. Other important issues are lemmatisation (lemma establishment, Svensén, 2009:94) and describing form variants and regional distribution of signs (McKee & McKee, 2013; Zwitserlood et al., 2013; Fenlon et al., 2015). Descriptions and decisions on these issues are based on the corpus data available[5].

In this process, corpus data have priority over additional data as it is usage data in comparison to elicited answers stemming from the DGS-Feedback. However, as we are dealing with a highly variable and non-standardised language (DGS) and as the corpus is relatively small – compared to the multi-million word corpora used for written language lexicography – it is helpful to have also other sources of information available when making lexicographic decisions. Data obtained by the DGS-Feedback adds information on the signs, supports the lexicographic work and therefore helps to improve the later product – the dictionary.

4. Contribution of DGS-Feedback Data

In the analysis stage of the lexicographic work the corpus data of one sign is analysed with regard to all dictionary-relevant facts, including meaning, form variation, regional distribution, and variation across different age groups. For all these facts corpus data may contain sufficient evidence to provide a clear-cut picture of the sign's properties and uses to be described. However, the corpus can only provide positive evidence of e.g. a variant form, a sense or regional distribution. Areas of uncertainty remain when there is only very little evidence in the corpus. Little or no data can either result from non-existence or from non-appearance of this feature in the corpus due to size, chance, and frequency of a sense. In these cases, additional data from the DGS-Feedback can be useful to obtain a clearer picture of the sign's properties. Furthermore it may add weight to the decision on which signs and meanings are to be selected for description in dictionary entries. The results of the corpus analysis are compared to the results from the DGS-Feedback to cross-check and supplement the findings. Doing so, we encounter different cases. The DGS-Feedback results can either confirm corpus data findings, or considerably differ from them. So far it does not seem useful to formulate strict guidelines or thresholds on how to weight *used* or *known* answers in comparison to corpus tokens, as all available information has to be taken into account to arrive at a comprehensive view on the sign's properties. DGS-Feedback results have to be interpreted carefully as a variety of factors can have influence on the outcome. These are e.g. the accidental participant sampling with respect to sociologic factors or the way question items are presented. In the following examples, we will discuss the most important ones.

[3] Within the charts representing DGS-Feedback results the different answers are represented in red (*used*), blue (*known*) and grey (*unknown*). Beige signifies areas where no participants contributed so far.

[4] iLex is the annotational and lexical database and working environment that is used for the DGS-Korpus project (Hanke & Storz, 2008).

[5] A more detailed description of the analysis of corpus data for lexicographic purposes are presented in Langer et al. (2018) in this issue.

4.1 Cases of Confirmation

DGS-Feedback data can confirm corpus findings in different respects – a good evidence of corpus tokens corresponds to many *used* responses, a scarce one to few positive responses.

4.1.1. Strong Corpus Evidence and High Positive DGS-Feedback Response

Strong corpus evidence alone would suffice for inclusion of a sense into the entry. If there are many *used* responses, DGS-Feedback results confirm this finding. This is the case for the form-meaning combination in example 1.

Example 1	
‘father’	
Sense	male parent, man who rears a child
Number of corpus tokens	156 from 63 informants
Total number of responses	147
Used	116
Known	27
Unknown	4

Table 1: ‘father’

4.1.2. Weak Corpus Evidence and Low Positive DGS-Feedback Response

If only few corpus tokens and a relatively low percentage of *used* or *known* answers are found, a closer look at the data is needed especially with regard to region, age, hearing status, and age of language acquisition, as these factors may have an influence on sign use of informants and response behaviour of DGS-Feedback participants.

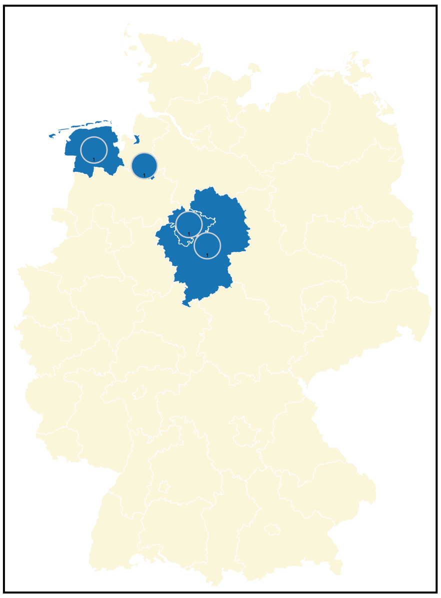


Figure 2: Distribution of corpus informants using ‘Monday’

In some cases low token numbers and a low DGS-Feedback percentage of *used* and *known* answers are both by themselves not conclusive while in combination can stabilise the findings and suggest an explanation. E.g. the low proportion of *used* in the case of ‘Monday’ (example 2) appears to be the result of a very regional distribution in Lower Saxony (see figure 2).

Example 2	
‘Monday’	
Sense	Monday, name of the first day of the week
Number of corpus tokens	9 from 4 informants
Total number of responses	104
Used (red, see figure 3)	5
Known (blue, see figure 3)	15
Unknown (grey, see figure 3)	84

Table 2: ‘Monday’

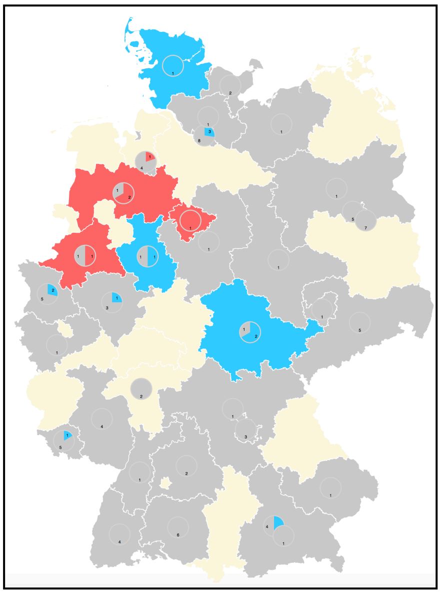


Figure 3: Distribution of DGS-Feedback responses concerning ‘Monday’

Example 2 is also a case where a more clear-cut picture of regional distribution results from the DGS-Feedback data. Although we recorded 330 different informants from different regions the information on regional signs is often rather scarce. Not every informant from a certain region uses every regional sign from his or her region. DGS-Feedback participants add here with their information on use and knowledge. In this case the majority of *used* or *known* responses either match with the region of Lower Saxony or come from participants living in adjacent parts of the country (see figure 3).

4.1.3. No Corpus Evidence and No or Low Positive DGS-Feedback Response

For some items there are no tokens in the corpus and also no or few *used* or *known* answers (see example 3). The core meanings of the sign in example 3 are 'food' and 'to eat', which are well attested. Another meaning is 'menu' in the sense of 'a list or range of food offered'. In spoken German the polysemous word *Menü,* which is the basis for a corresponding mouthing, may also denote the menu options of computer programs. Because of this, the sign from example 3 could possibly be used to express 'menu (computer)' although this meaning is not related to food. In this case, the DGS-Feedback data supports the impression from the corpus that this sense is very likely not an established use in the sign language community. Unless further evidence emerges such a sense will not be included in the dictionary entry.

Example 3	
'menu'	
Sense	small display on the computer to choose editing options
Number of corpus tokens	0
Total number of responses	103Beispiele
Used	3
Known	14
Unknown	86

Table 3: 'menu (computer)'

Example 3 is a result of the workflow established to verify or disprove data from previously published sources. Here all listed and presumed form-meaning combinations of the sign in question have been put into the DGS-Feedback for verification independently of corpus evidence (Langer et al., 2014).[6] Many of the previously published sources are sign collections that are based on German wordlists (see e.g. Johnston 2003 for a critical view on publications based on wordlists). Asking signers for their sign equivalents for words off a word list is a method that elicits not only established signs. It is also prone to produce some spontaneous isolated sign uses that are not actually established in the signing community. Especially when the items concern technical terms or new concepts that may not have established signs yet. Some of these artefacts have made their way into sign collections. This might also apply to example 3 taken from *Fachgebärdenlexikon Computer* (Arbeitsgruppe Fachgebärdenlexika, 1994). Findings like example 3 show that the DGS-Feedback can be useful in filtering out such artefacts.

[6] At the present stage of the project lexicographic descriptions are fully based on corpus evidence. That means the DGS-Feedback now is only used to check meanings of low token evidence but not items that have no token evidence at all.

4.2 Cases of Discrepancy

In some cases corpus and DGS-Feedback results differ considerably from each other. These cases require a closer look and ask for an explanation.

4.2.1 Strong Corpus Evidence and Low Positive DGS-Feedback Response

If there is a high token number for a certain meaning but little *used* or *known* answers in the DGS-Feedback, it would still be included as a sense in the dictionary, because corpus data has priority over the DGS-Feedback data. However, we always try to find a plausible explanation for discrepancies in the two data sources. For example, they may be a result of differences in sampling as in the following example 4. While there is good corpus evidence for the sign of example 4 to have the meaning of 'bread', DGS-Feedback responses do not seem to confirm this finding.

Example 4	
'(loaf of) bread'	
Sense	food made of flour, water and yeast
Number of corpus tokens ("non-tokens" excluded)	26 from 16 informants
Total number of responses	71
Used (red, see figure 5)	3
Known (blue, see figure 5)	14
Unknown (grey, see figure 5)	54

Table 4: '(loaf of) bread'

The relatively high token count for this meaning is a result from a particular elicitation task. With this task signs for certain concepts (e.g. bread) known to be highly variable from region to region were elicited.[7] It was to be expected that otherwise findings of such regional signs, that we want to document, would be scarce. But, even though the majority of tokens (19) appear in the context of this task, there are also findings of the sign (7) within tasks that have conversational character. In the corpus data, regional

[7] Only in one of the 20 tasks in the corpus elicitation the participants were directly asked to show their sign for a given concept and to give an example sentence. All other tasks used within the DGS-Korpus project aimed at more natural signing or for free conversational data (Nishio et al., 2010). A direct elicitation of this kind produces metalinguistically aware answers as opposed to spontaneous sign use. Informants often do not only show their own sign but also other signs they know for the concept, which should not be counted as an evidence of their personal sign use. This problem was addressed in the paper on so-called "non-tokens" (cf. Langer et al., 2016b).

distribution as a variant for ‘bread’ in the Bavarian and Hessian area is well evidenced (see figure 4).

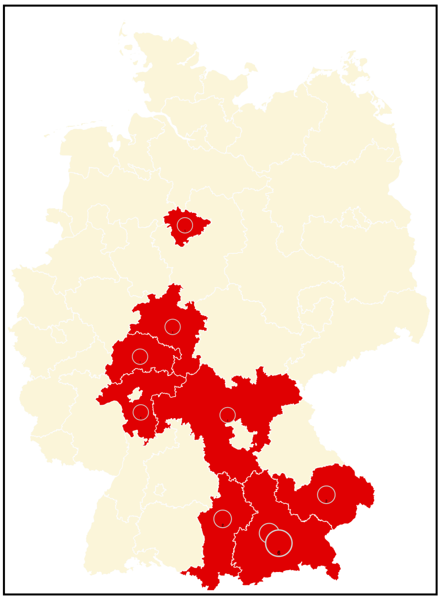


Figure 4: Distribution of corpus informants using ‘bread’

In the DGS-Feedback data, 3 of 8 of the participants from Bavaria answered with *used* and further 4 answered with *known*. Up to date only one user from the Hessian region participated and voted *unknown* (see figure 5). Altogether only few DGS-Feedback participants were from the area of sign use that is evidenced from the corpus data.

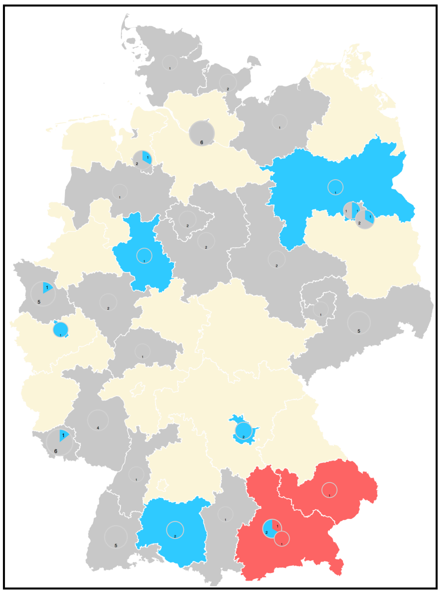


Figure 5: Distribution of DGS-Feedback responses concerning ‘bread’

Taking the information from corpus and DGS-Feedback data together a rather restricted region (see figure 6) becomes apparent. Most tokens and *used* answers (orange in figure 6) stem from Southern Bavaria indicating that the sign is mainly used in that area. For the dictionary this would mean a note on regionality.

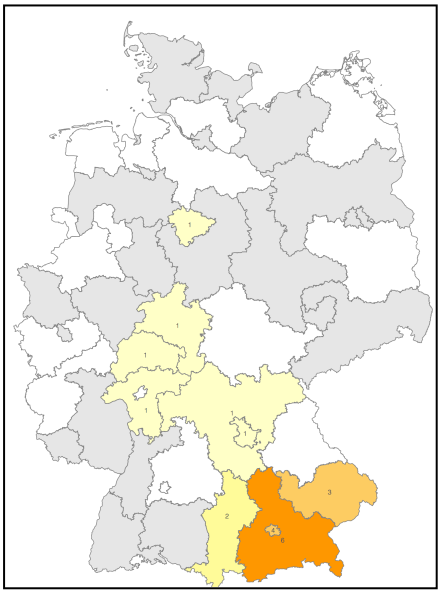


Figure 6: Distribution of corpus informants and DGS-Feedback participants using ‘bread’

4.2.2 Weak Corpus Evidence and High Positive DGS-Feedback Response

Example 5	
‘earring’	
Sense	jewellery worn on the ear
Number of corpus tokens	6 from 4 informants
Total number of responses	139
Used	121
Known	14
Unknown	4

Table 5: ‘earring’

In some cases there are only few corpus occurrences but the percentage of *used* answers is high. It is reasonable to assume that corpus evidence is low because the sign or sign sense is a low-frequency item, or because it is not appearing in the corpus very often as no relevant topic has come up during elicitation sessions, or because the sign with this sense is rarely used in communicative events as recorded. In this case the DGS-Feedback provides us with a good reason to include a sense into an entry. Otherwise it would have been held back until token count for the sense would have risen.[8] In a case like this the

[8] Senses that have only weak corpus evidence are nevertheless documented in the internal pre-dictionary database and put to the status of *under surveillance*. As corpus annotation is ongoing further corpus evidence may emerge. Items *under surveillance* will not appear in the dictionary entry at the current state but

DGS-Feedback results provide an additional basis for decision-making. An example for such a case is the sense ‘earring’ (example 5). The iconic value of the sign is a representation of a ring or bud in the ear. This sign may be used for ‘earring’ as well as for the well-evidenced senses ‘woman’ or ‘girl’. So, even if the sense ‘earring’ is not well represented in the corpus, the DGS-Feedback gives a good reason to include the sense as many *used* answers indicate it as a conventional meaning of that sign.

4.2.3 No Corpus Evidence and High Positive DGS-Feedback Response

In the last case to be discussed no corpus evidence for a sense could be found but in the DGS-Feedback there was a high percentage of *used* answers. This leads to a preliminary description of this sense within the pre-dictionary database, but with the status *under surveillance*. We prefer corpus evidence over DGS-Feedback data as the goal is to produce a corpus-based dictionary. Additionally, senses are usually illustrated by examples taken from the corpus. So senses without corpus evidence will not be included into the product until there is at least some evidence from corpus data.

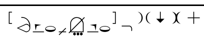
Example 6	
‘medical’ 	
Sense	of a or concerning a doctor
Number of corpus tokens	0
Total number of responses	124
Used	87
Known	19
Unknown	18

Table 6: ‘medical’

Even though *used* answers are high for example 6 other factors need to be considered. It is not always easy to create good stimuli for the surveys, especially if we try to verify or disprove meanings expressed by German words (translational equivalents) stemming from word lists of sign collections. Transferring a sense like *ärztlich* (‘medical’, see example 6) into a signed context is not easy. Knowledge of German and the presented translational equivalents can have an influence on the responsive behaviour of the participants. Thus an overall acceptance of a certain form-meaning combination is possible if the German word is known, even though the concept might usually be expressed differently within the community. So language contact might play a role here.

may be included in the future if sufficient evidence can be found.

4.3 Participant Comments on Sign Use

Participants are given the option to comment on sign use. These comments give interesting insight into homonyms, additional senses, further form variation, lexical variants, and problems of understanding concerning the presented stimulus. Such information is valuable for the dictionary writing as well as for the enrichment of the lexical database. Signs having same or similar forms are cross-referenced in the lexical database and in the dictionary entries. Comments from the DGS-Feedback provide hints on such relations between signs that have been missed so far.

Example 7	
Sign: 	
Core sense	‘eye’
From related sign: 	
Core sense	‘to try’
Number of written comments	11

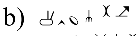
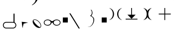
Table 7: Form-related signs

Example 7 shows such a finding that resulted from the comments that were given on the sign form with the core sense ‘eye’ at the end of the question concerning that form. There were 11 written and 1 signed answer(s) that this sign could also mean ‘to try’. In iLex we have two well evidenced sign type entries for ‘eye’ and ‘to try’ showing slightly different citation forms. The only difference is the location of the sign. The form with the core meaning ‘eye’ is usually signed close to the eye at the upper part of the cheek. In comparison, the sign with the core meaning ‘to try’ is signed at the cheek but not necessarily close to the eye. Both signs are so similar in form that, when presented in isolation, they could be mistaken for each other. Following the comments of the participants a new cross-reference was added in the annotational database for these two signs.[9]

Cross-references within the database that can be established through these findings are beneficial for transcription, as they help annotators to find signs within

[9] The location of a body-anchored sign in actual use may be within a more ore less extended area of contact rather than only one specific spot. Areas of different signs with different locations can be overlapping. For the purpose of quick type identification in the database a citation form of each sign type is defined by a HamNoSys Notation (Hanke, 2004). When working on an entry the review of token data can lead to a correction of the citation form. When establishing lemmas it has to be checked whether the two type entries in the annotational database ‘eye’ and ‘try’ have to be merged into one dictionary entry or whether they are better described in two separate entries (cf. Langer et al., 2016c). Cross-references in the database support this step by bringing sign types that are similar to the respective entry candidate to the notice of the lexicographer and making them easily accessible in the database for inspection.

the database more easily. Additionally the dictionary entries profit from this information as cross-references to similar signs are included in the entries of the future dictionary.

Some participants also use the video function to show their sign for a meaning. This is usually the case when a presented form-meaning combination is not accepted by them. For example within the DGS-Feedback questions the sign a)  with the meaning ‘to watch out’ was asked for. Within the video comments two participants answered that they use sign b)  to express ‘to watch out’ instead of sign a) .

In some cases it makes sense to conduct a spot transcription. Such transcribed video answers supplement the corpus findings. So when WSD for the sign form starts these “tokens” are available in the database and may be consulted in addition to the corpus findings.

5. Conclusion

Data from the DGS-Feedback adds valuable information on the signs, their forms and meanings in addition to the findings from the corpus. It can confirm uncertain sign use, help to find special characteristics of signs (e.g. regional use, form variation, age effects) and can be utilized to improve the content of the annotation database.

Up to now DGS-Feedback data has been collected with question types targeting basic vocabulary. To suit the needs of the corpus-based WSD and dictionary writing process better, new questions types for the DGS-Feedback System will be developed. One question type already in preparation focuses on specific senses that have only very weak corpus evidence. This means that evidence is not stable enough to base a well-informed decision on inclusion or exclusion of the sense into the entry or not on the grounds of corpus data alone. Thus supplementing data from the DGS-Feedback may be helpful here.

In general, the data from the DGS-Feedback System need to be analysed and interpreted carefully when compared to the corpus findings especially if they seem to differ from the corpus evidence. As we have shown in the examples 2 and 4 to 6, there is no reliance on numbers of response alone. However, in combination with corpus evidence they often are helpful in lexicographic decision-making.

6. Acknowledgements

This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies’ Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies’ Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

7. Bibliographical References

Atkins, B.T.S., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography.* Oxford: Oxford University Press.

Fenlon, J., Cormier, K., Schembri, A. (2015). Building BSL SignBank: The Lemma Dilemma Revisited. *International Journal of Lexicography, Volume 28, Issue 2*. Oxford: Oxford University Press, pp. 169--206.

Hanke, T., Hong, S.-E., König, S., Langer, G., Nishio, R., Rathmann, C. (2010). Designing Elicitation Stimuli and Tasks for the DGS Corpus Project. Poster presented at the Theoretical Issues in Sign Language Research Conference (TISLR 10), Sept 30 - Oct 2, 2010 at Purdue University, Indiana, USA.

Hanke, T., Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, & E. Thoutenhoofd (Eds.), *Construction and Exploitation of Sign Language Corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages. 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco*. Paris: ELRA, pp. 64--67.

Hanke, T. (2004). HamNoSys – representing sign language data in language resources and language processing contexts. In O. Streiter & C. Vettori (Eds.), *From SignWriting to Image Processing. Information techniques and their implications for teaching, documentation and communication. Proceedings of the Workshop on the Representation and Processing of Sign Languages. 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal*. Paris : ELRA, pp. 1--6.

Johnston, T. (2003). Language Standardization and Signed Language Dictionaries. *Sign Language Studies, 3.4.* Washington D.C.: Gallaudet University Press, pp. 431--468.

König, S., Matthes, S., Barbeito Rey-Geißler, P., Blanck, D., Hanke, T., Konrad, R., Langer, G. (2013). Beteiligung der Sprachgemeinschaft. DGS-Korpus – Feedback-System. Poster zur 5. Nacht des Wissens am 02.11.2013 an der Universität Hamburg.

Langer, G., Müller, A., Wähl, S. (2018). Queries and Views in iLex to Support Corpus-based Lexicographic Work on German Sign Language (DGS). In this issue.

Langer, G., König, S., Matthes, S., Groß, N., Hanke, T. (2016a). Variation of DGS lexical items. What sign language lexicography can gain from a mixed method approach: Corpus data supplemented by crowd sourcing. Poster presented at the international Conference on Theoretical Issues in Sign Language Research (TISLR 12), Jan 4-7, 2016 at Melbourne, Australia.

Langer, G., Hanke, T., Konrad, R., König, S. (2016b). “Non-tokens”: When Tokens Should not Count as Evidence of Sign Use. In E. Efthimiou, E. Fotinea, T. Hanke, J. Hochgesang, J. Kristoffersen, & J. Mesch (Eds.), *Corpus Mining. Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages. 10th International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia.* Paris: ELRA, pp. 137--142.

Langer, G., Troelsgård, T., Kristoffersen, J., Konrad, R., Hanke, T., König, S. (2016c). Designing a Lexical Database for a Combined Use of Corpus Annotation and Dictionary Editing. In E. Efthimiou, E. Fotinea, T. Hanke, J. Hochgesang, J. Kristoffersen, & J. Mesch (Eds.), *Corpus Mining. Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages. 10th International Conference on*

Language Resources and Evaluation (LREC 2016), Portorož, Slovenia. Paris: ELRA, pp. 143--152.

Langer, G., König, S., Matthes, S. (2014). Compiling a Basic Vocabulary for German Sign Language (DGS) – lexicographic issues with a focus on word senses. In A. Abel, C. Vettori, & N. Ralli (Eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus, July 15-19 2014 in Bolzano/Bozen – Italy.* Bolzano/Bozen: EURAC research, pp. 767--786.

Nishio, R., Hong, S.-E., König, S., Konrad, R., Langer, G., Hanke, T., Rathmann, C. (2010). Elicitation methods in the DGS (German Sign Language) Corpus Project. In P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Martínez Ruiz, & A. Schembri (Eds.), *Corpora and Sign Language Technologies. Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages. 7th International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta.* Paris: ELRA, pp. 178--185.

McKee, R., D. McKee. (2013). Making an Online Dictionary of New Zealand Sign Language. *Lexikos, 23.1*. Stellenbosch: Buro van die Wat, pp. 500--531.

Svensén, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making.* Cambridge: Cambridge University Press.

Arbeitsgruppe Fachgebärdenlexika (Eds.). (1994). *Fachgebärdenlexikon Computer*, vol. 2. Hamburg: Signum Verlag.

Zwitserlood, I., Kristoffersen, J., Troelsgård, T. (2013). *Issues in Sign Language Lexicography.* In Jackson, H. (Ed.), *The Bloomsbury Companion to Lexicography.* London: Bloomsbury Publishing, pp. 259--283.

Exploring Localization for Mouthings in Sign Language Avatars

Rosalee Wolfe[1], Thomas Hanke[2], Gabriele Langer[2], Elena Jahn[2], Satu Worseck[2], Julian Bleicken[2], John C. McDonald[1], Sarah Johnson[1]

[1]School of Computing, DePaul University, 243 S. Wabash Ave., Chicago IL 60604, USA
[2]Institute of German Sign Language and Communication of the Deaf (IDGS), University of Hamburg, Gorch-Fock-Wall 7, 20354 Hamburg, Germany

rwolfe@depaul.edu
{thomas.hanke,gabriel.langer,elena.jahn,satu.worseck,julia.bleicken}@uni-hamburg.de
jmcdonald@cs.depaul.edu, sarahej101@gmail.com

Abstract

According to the World Wide Web Consortium (W3C), localization is "the adaptation of a product, application or document content to meet the language, cultural and other requirements of a specific target market" (Ishida & Miller, 2015). One aspect of localizing a sign language avatar is creating a capability to produce convincing *mouthing.* For purposes of this inquiry we make a distinction between *mouthings* and *mouth gesture* (Crasborn et al., 2008). The term 'mouthings' refers to mouth movements derived from words of a spoken language while 'mouth gesture' refers to mouth movements not derived from a spoken language. This paper reports on a first step to identify the requirements for an avatar to be capable of mouthings in multiple signed languages.

Keywords: avatar technology, mouthing, German Sign Language (DGS), lip sync

1. Mouthings in signed languages

The occurrence of mouthings has been reported for many signed languages (cf. for example (Boyes-Braem & Sutton-Spence, 2001) and their origin in spoken languages are self-evident. The prevalence of mouthings varies across different sign languages and individual signers. In German Sign Language (DGS) the occurrence of mouthings is very common. In (Ebbinghaus & Heßmann, 1996) researchers report that the mouthings may be changed and adapted when used in signing. Examples of this include dropping grammatical endings, exaggerating selected elements, holding end position as in the case of word-final L in Apfel or adapting rhythm of mouthed syllables to the rhythm of the manual singing.

While mouthings occur regularly in most sign languages, their significance and status have been a matter of sometimes heated discussions among sign linguists. For example, there is no consensus on the role of mouthings in American Sign Language (ASL) (Lucas & Valli, 1989; Nadolske & Rosenstock, 2007).

However, no matter the theoretical viewpoint one takes on the issue of mouthing, one must acknowledge that for most if not all sign languages mouthings do occur. If an avatar purports to fully express any signed language, it must have the capacity to express all aspects of the language which likely will include mouthings. Without mouthings, avatar signing would not only look unnatural for most sign languages and could also omit important information, resulting in utterances that could be incomprehensible. However, the avatar should also have sufficient flexibility to omit mouthings all together and limit its production exclusively to mouth gestures.

2. History of lip sync technology

Portraying mouthings requires animating an avatar's mouth. Animating an avatar's mouth originates with the technique of creating speaking characters. These first appeared with the advent of sound cartoons in the 1920s (Fleischer, 2005). A believable speaking character requires lip motion that moves in synchrony with a pre-recorded sound track (Johnson & Thomas, 1995), hence the name *lip sync*. Animators drew images to portray *visemes,* or the shape that the lips take while producing the phonemes of a spoken dialog (Fisher, 1968). Because some phonemes appear identical on the face even though they have different sounds, lip sync requires fewer visemes than there are phonemes in a language. For example, for lip sync of English dialog, animation artists typically use between seven and 12 visemes to represent the 44 phonemes of the spoken language (Halas & Manvell, 1971 ; Johnson & Thomas, 1995). However, for extremely simple animation, animators reduce this number to four (Atkinson, 2017) or even two (Hess, 2016). This was a manual, time-consuming process, with the artist being responsible for viseme selection and timing.

The turn of the century witnessed the rise of multimodal technology, which integrated audio and video output in intelligent agents for an enhanced interactive user experiences (Kshirsagar et al., 2002). The intelligent agents were embodied as *avatars* which are representations of human figures. To enhance its human-like qualities, the avatar must move its lips in synchrony with its speech output. This requires the automation of the lip sync animation.

Similar to manually-produced animation, automated lip sync requires a sound track and visemes to generate a talking figure, but it differs in its representation of visemes. The automation strategies fall into two categories, based on the avatar's representation of visemes. In video-based 2D approaches, computer vision techniques analyze frames of pre-existing video recordings and extract the visemes as groups of pixels (Theobald et al., 2003). To accommodate a new sound track, the software identifies and changes the visemes in existing frames of the video to synchronize the lips with the new phoneme stream. When dubbing a movie in a foreign (spoken) language, this technique helps with synchronizing an actor's lip movements with the translated dialog.

In synthetic 3D approaches, the visemes are not sets of pixels, but collections of 3D data. An avatar system can rely directly on a set of artist-created models of an avatar's lip positions to depict visemes. These can use *blend shapes* expressed as polygon meshes, or they can utilize a muscle-based system (Deng & Noh, 2008). Alternatively, it can utilize a high-level animation standard such as MPEG-4 Face and Body Animation which consists of a set of predefined Facial Animation Parameters (FAPs) including 14 static visemes (Pandzic & Forchheimer, 2003).

In 3D strategies, the technique used to generate the animation depends on the source of the dialog. In the case where there is a prerecorded voice track, a speech-recognition module can detect the phonemes and select the corresponding viseme (Zorić & Pandžić, 2005). The viseme choice and timing become part of the data that the animation system interpolates to create the individual frames of the animation. In the case where there is no prerecorded voice track, but only a text containing the dialog, this approach can still be effective if there is a text-to-speech (TTS) service available. Many TTS services provide an option to produce phonemes and timing information as text, which can easily be converted into a stream of viseme choices with timing.

No matter the strategy, there is a question of how best to choose the visemes to match the spoken phonemes for automatic lip sync. (Chen & Rao, 1998) suggested that the possibility of using data-analysis techniques to analyze video recording with the goal of identifying the visemes. However, (Cappelletta & Harte, 2012) examined five phoneme-to-viseme mappings for visual speech recognition, four of which were developed through data analysis and one which was created by linguists. They found that the linguistically-motivated viseme mapping performed the best on visual-only recognition of continuous speech.

3. Lip synch technology for enhanced accessibility

Although most interactive lip sync systems were created for hearing communities, several technologies emerged to improve speech recognition for those who are hard-of-hearing or who find themselves in noisy environments. An early example was a multimedia telephone to assist the hard-of-hearing which used a simple "2-1/2D" head that portrayed lip sync to accompany the voice data (Lavagetto, 1995). A similar project (Oviatt & Cohen, 2000) strove to enhance speech recognition for hearing people located in noisy environments.

One of the distinguishing characteristics between a person who is hard-of-hearing and a person who is Deaf is their language preference. A person who is hard-of-hearing prefers a spoken language, but will use assistive technology such as hearing aids or closed captioning to gain better access to the content of spoken messages. In contrast a person who identifies as Deaf will use a signed language, such as ASL or DGS as their preferred language (Padden & Humphries, 1988).

For the Deaf community, access to spoken language, or to the written form of a spoken language requires translation to the preferred signed language. An essential part of any automatic spoken-to-sign translation system is an avatar capable of producing all aspects of the language, including mouthings. The earliest avatar designed specifically to provide improved access to the Deaf community was part of the ViSiCAST project (Elliott et al., 2000). This project included the development of the Signing Gesture Markup Language (SiGML), based on HamNoSys (Hanke, 2004). It specifies a *mouth picture* or viseme for each letter of the International Phonetic Alphabet (IPA) (Glauert et al., 2004). Strings of visemes are expressed using SAMPA encoding conventions and the mapping of SAMPA symbols to visemes is part of an avatar-specific configuration file. The mapping was subsequently revised, and the current pronunciation dictionary used for DGS is from IKP Bonn (Aschenberner & Weiss, 2005).

The sign annotation software iLex uses the same system for annotating mouthings in lexical items (Hanke, 2002). (Elliott et al., 2008) describe the continuation of this research as part of the eSIGN project, and gives a complete example of SiGML notation, including mouthings, for the DGS sign HAUS, as well as selected frames from an avatar signing HAUS. (Jennings, Elliott, Kennaway, & Glauert, 2010) give an in-depth discussion of the implementation details for Animgen, the animation engine used to create the avatar. To implement mouthings, they use a set of blend shapes, one for each mouth picture.

Contemporaneous with the ViSiCAST/eSIGN projects, other groups explored the possibility of incorporating mouthings in sign language technology. These include projects at the German Research Center for Artificial Intelligence (DFKI) (Heloir, Nguyen, & Kipp, 2011) (Kipp, Heloir, & Nguyen, 2011) and DePaul University (Wolfe et al., 2009). The primary goal of the avatar developed at DFKI is to synthesize DGS, and it uses the OpenMARY speech synthesis system to generate the viseme specification and timing, but no mention was made of the underlying technology for representing individual visemes. In contrast, the avatar "Paula" developed at DePaul generates ASL, and uses a Microsoft.NET Text-to-Speech (TTS) service to generate the viseme selection and timing. Because the face is represented by a muscle system, Paula's mouth animation is not limited to linear combinations of selected visemes.

4. Extending a muscle-based avatar

In a past study (Schnepp, Wolfe, McDonald, & Toro, 2012), members of the Deaf community in the US viewed the Paula avatar with and without mouthings and consistently indicated a preference for animations with mouthings. Encouraged by this feedback, we are exploring the feasibility of adding localization to Paula. As a first step we attempted to teach Paula to sign DGS, for which mouthings are an important feature. For this first inquiry, we chose six signs from a previously existing vocabulary for Swiss German Sign Language (Ebling, et al., 2017) whose manual channel match signs in DGS. See Figure 1.

Creating the mouthings posed several challenges. The TTS library was specific to English, and had occasional difficulties in synthesizing spoken German words. Correcting these instances required manual editing of several of the generated viseme streams.

APFEL
BALL
BÄR
JUNGE
UHR
ZEIT

Figure 1: DGS signs under investigation

Another challenge was the style of enunciation. In general, native US speakers of English demonstrate an economy of lip motion in conversation. This, coupled with the lack of consensus on the role of mouthings in ASL lead to a previous design decision to keep Paula's lip movement to a minimum and to provide an option to omit it all together.

In contrast, mouthings often appear in DGS. Furthermore, there are important differences between German and English bases of articulation (Hall, 2003). Spoken German has a greater articulatory tension; muscles in the articulators such as the tongue are tenser, resulting in pronunciations that are more forceful. The tongue takes on positions that are more extreme and more prominent. Lip movements are more vigorous in German. Vowels such as /u:/ and /y:/ are articulated with strongly protruding and rounded lips.

These differences in the basis of articulation required adjustment of the viseme weights. Instead of the standard 30% of maximum viseme strength that had been used to accompany ASL, the DGS settings ranged from 50% to 120% of the (original) maximum. Figure 1 demonstrates the difference in the spoken English viseme and DGS mouth shape for the /s/ phoneme. We were motivated by the feedback of one of our German colleagues, who said, "I need to see more teeth!"

US-English viseme | Prototype DGS mouth shape

Figure 2: Mouth shapes for /s/

Informed by data from the DGS-Korpus (Blanck, et al., 2010) we also adjusted the timing of the viseme onset. Instead of coinciding with the onset of the manual channel of a lexical item, the mouthings in DGS tend to start earlier. Based on this finding, we set the onset of the lip motion to begin 0.2 seconds before the onset of the manual channel.

5. A first feedback session

A group of six linguistically aware native signers of the German Deaf community participated in a first feedback session. The session began with a brief introduction to avatar technology and its possible applications. Then, to familiarize the group with the current capabilities of avatar technology, a moderator presented three short animations that demonstrate the state of the art in sign language technology (Jordaan, 2014 ; Brun, 2014 ; The ASL Avatar Project Team at DePaul University, 2012), and conducted a discussion that compared the three animations.

A second moderator presented the newly-created DGS signs complete with a typical mouthing. Each sign was presented as a series of three slides. See Figure 3. The first slide simply gave an identification number for the sign. The next two slides contained the same identification number and a video frame. The first video used a medium shot showing the avatar from the waist up. The second video showed the same sign, but used a close-up shot, to show the mouth in extreme detail. The moderator played the videos as many times as group members requested.

After playing the first (medium shot) video, the moderator asked the group to identify the sign, and solicited comments on what they liked and what needed to be improved. After playing the second (close-up shot) video, the moderator again solicited comments on what needed improvement.

Other than the brief introduction, which was presented in written German and interpreted into DGS, the entire session was conducted by two Deaf moderators. To accommodate the hearing note takers, the discussions were voiced in German by an interpreter.

6. Feedback

In all cases, the signs were immediately identified, which was consistent with results previously received from a focus group fluent in DSGS (Ebling, et al., 2017). The color selections for the avatar clothing, hair and background made it easy to read the manual channel and it was well positioned in the signing space. However, the lighting on the face was too even and needs to reveal the contours of the lower face. Viewers wanted to the nasolabial folds (smile lines) to be clearly visible at all times.

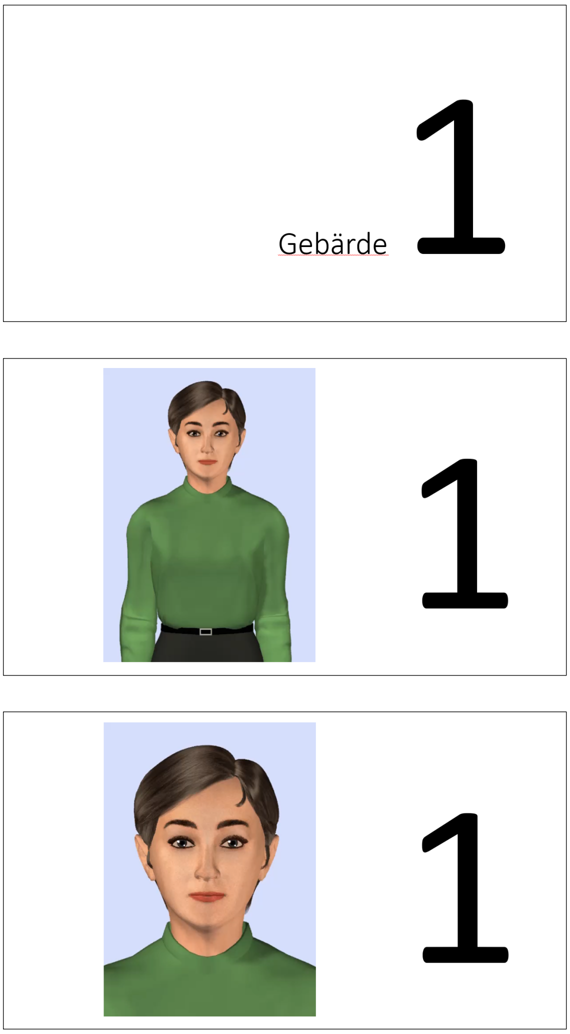


Figure 3: Slide format for presenting avatar signing DGS

There were several issues identified with the mouthings. There was general agreement that all of the visemes need to be more "pronounced". The teeth needed to be more prominent. The word-final viseme corresponding to /l/ in BALL and APFEL requires the tongue to be farther forward in the mouth, with the blade of the tongue at the alveolar ridge, and the tongue tip behind the upper teeth. This is consistent with the findings of (Ebbinghaus & Heßmann, 1994).

However, there was one aspect of creating a more vigorous pronunciation that had nothing to do with the lips themselves. Group members consistently pointed to a lack of cheek motion in the mouthings. They indicated that cheek movement is important for all visemes, but are particularly vital for the labial plosives /b/ and /p/, and demonstrated how the cheek movement is necessary for a mouthing that is easy to recognize.

7. Conclusions and future work

In this effort, we explored the viability of avatar localization, to identify challenges of adapting an avatar to produce sign languages from different geographic regions. We started with an avatar that was designed to produce ASL, and used it to create lexical items in DGS. The major change was to modify its capabilities to produce DGS mouthings. We then solicited feedback from linguistically aware native DGS signers. Although all of the lexical items were immediately identified, there were issues with several of the visemes comprising the mouthing.

Previously, focus of viseme development has been almost exclusively on mouth shape. Future avatar development will need to consider how to incorporate areas surrounding the mouth including cheeks and nose for improved legibility.

Our preliminary findings seem to run counter to (Glauert 2004)'s supposition that if single set of visemes will suffice for mouthings in all signed language. The visemes we created to support spoken English are inadequate for DGS. It will be necessary to create a library of visemes, preferably by artists aware of the role of mouthings in DGS. However, for effective production of mouthings it will not suffice to use such a library with a simple surface mapping from audible phonemes produced by a TTS. Ultimately, it will require corpus data that contain instances consistent with the findings of (Elliott E. A., 2013) and (Ebbinghaus & Heßmann, Signs and words: Accounting for spoken language elements in German Sign Language, 1996) that demonstrate the adaptation of mouthings as produced in DGS.

However, it would be interesting to further explore the possibility of viseme reuse to support mouthings for multiple signed languages. Creating a set of language-independent visemes for spoken languages has been a topic of research for some time (Zorić & Pandžić, Real-time language independent lip synchronization method using a genetic algorithm, 2006). However, attempting to extend this idea to signed languages is an open question and an intriguing topic for future work.

8. Acknowledgments

This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

9. References

Aschenberner, B., & Weiss, C. (2005). *Phoneme-Viseme Mapping for German Video-Realistic Audio-Visual-Speech-Synthesis IKP-Working Paper NF 11.* Retrieved 2018, from CiteseerX: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.8839&rep=rep1&type=pdf

Atkinson, D. (2017, December 24). *Frank Hellard's mouth shapes.* Retrieved December 24, 2017, from Centre for Animation & Interactive Media, RMIT University: http://minyos.its.rmit.edu.au/aim/a_notes/mouth_shapes_02.html

Blanck, D., Hanke, T., Hofmann, I., Hong, S.-E., Jeziorski, O., Kleyboldt, T., . . . Wager, S. (2010). The DGS Corpus Project. Development of a Corpus Based Electronic Dictionary German Sign Language – German. *Theoretical Issues in Sign Language Research (TISLR) 10.* West Lafayette, Indiana.

Boyes-Braem, P., & Sutton-Spence, R. (2001). *The hands are the head of the mouth.* Hamburg: Signum-Verlag.

Brun, R. (2014, May 01). *Sign Language: Gallaudet University | Mocaplab.* Retrieved December 24, 2017, from mocaplab: http://www.mocaplab.com/projects/gallaudet-university/

Cappelletta, L., & Harte, N. (2012). Phoneme-to-viseme Mapping for Visual Speech Recognition. *ICPRAM (2)*, (pp. 322-329).

Chen, T., & Rao, R. R. (1998). Audio-visual integration in multimodal communication. *Proceedings of the IEEE, 86*, 837-852.

Crasborn, O. A., Van Der Kooij, E., Waters, D., Woll, B., & Mesch, J. (2008). Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. *Sign Language \& Linguistics, 11*, 45-67.

Deng, Z., & Noh, J. (2008). Computer facial animation: A survey. In *Data-driven 3D facial animation* (pp. 1-28). Springer.

Ebbinghaus, H., & Heßmann, J. (1994). Formen und Funktionen von Ablesewörtern in gebärdensprachlichen Äußerungen (Teil I). *Das Zeichen, 30*, 480-487.

Ebbinghaus, H., & Heßmann, J. (1996). Signs and words: Accounting for spoken language elements in German Sign Language. (R. Wilbur, Ed.) *International review of sign linguistics, 1*(1), 23-56.

Ebling, S., Johnson, S., Wolfe, R., Moncrief, R., McDonald, J., Baowidan, S., . . . Tissi, K. (2017). Evaluation of Animated Swiss German Sign Language Fingerspelling Sequences and Signs. *Universal Access in Human-Computer Interaction*, 3-13.

Elliott, E. A. (2013). Phonological Functions of Facial Movements: Evidence from deaf users of German Sign Language. Berlin: Doctoral dissertation, Freie Universität Berlin.

Elliott, R., Glauert, J. R., Kennaway, J. R., & Marshall, I. (2000). The development of language processing support for the ViSiCAST project. *Proceedings of the fourth international ACM conference on Assistive technologies*, (pp. 101-108).

Elliott, R., Glauert, J. R., Kennaway, J. R., Marshall, I., & Safar, E. (2008). Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society, 6*, 375-391.

Fisher, C. (1968, December). Confusions among visually perceived consonants. *Journal of Speech and Hearing Research, 1*(11), 796-804.

Fleischer, R. (2005). *Out of the Inkwell: Max Fleischer and the Animation Revolution.* Lexington: University Press of Kentucky.

Glauert, J. R., Kennaway, J. R., Elliott, R., & Theobald, B.-J. (2004). Virtual human signing as expressive animation. *Symposium on Language, Speech and Gesture for Expressive Characters, University of Leeds*, (pp. 98-106).

Halas, J., & Manvell, R. (1971). *The Technique of Film Animation.* Waltham: Focal Press.

Hall, C. (2003). *Modern German pronunciation: An introduction for speakers of English.* Manchester University Press.

Hanke, T. (2002). iLex-A tool for Sign Language Lexicography and Corpus Analysis. *LREC.* Las Palmas: ELRA.

Hanke, T. (2004). HamNoSys -- Representing sign language data in language resources and language processing contexts. *Fourth International Conference on Language Resources and Evaluation (LREC 2004). Representation and Processing of Sign Languages Workshop* (pp. 1-6). Paris: European Language Resources Association.

Heloir, A., Nguyen, Q., & Kipp, M. (2011). Signing avatars: A feasibility study. *The Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT), Dundee, Scotland, United Kingdom.*

Hess, D. (2016, August 17). *Lip Semi-Sync: The Incomplete Guide to Mouth Flapping.* Retrieved December 24, 2017, from DIY Animation Club: https://diyanimation.club/animation-lip-sync-mouth-flapping/

Ishida, R., & Miller, S. K. (2015, April 9). *Localization vs. Internationalization.* Retrieved December 23, 2017, from World Wide Web Consortium: https://www.w3.org/International/questions/qa-i18n

Jennings, V., Elliott, R., Kennaway, R., & Glauert, J. (2010). Requirements for a signing avatar. *Proceedings of the 4th LREC Workshop on the Representation and Processing of Sign Languages*, (pp. 133-136).

Johnson, O., & Thomas, F. (1995). *The Illusion of Life: Disney Animation.* New York: Random House (Disney Press).

Jordaan, B. (2014, November 13). *Introductory video of myself at "Festmeny" show.* Retrieved December 24, 2017, from Facebook: https://www.facebook.com/braam3D/videos/10153283146689409/

Kipp, M., Heloir, A., & Nguyen, Q. (2011). Sign language avatars: Animation and comprehensibility. *Intelligent Virtual Agents*, (pp. 113-126).

Kshirsagar, S., Magnenat-Thalmann, N., Guye-Vuilleme, A., Thalmann, D., Kamyab, K., & Mamdani, E. (2002). Avatar markup language. *ACM International Conference Proceeding Series*, *23*, pp. 169-177.

Lavagetto, F. (1995). Converting speech into lip movements: A multimedia telephone for hard of hearing people. *IEEE Transactions on Rehabilitation Engineering, 3*, 90-102.

Lucas, C., & Valli, C. (1989). Language contact in the American deaf community. In C. Lucas (Ed.), *The Sociolinguistics of the Deaf Community* (pp. 11-40). San Diego, CA: Academic Press.

Nadolske, M. A., & Rosenstock, R. (2007). Occurrence of mouthings in American Sign Language: A preliminary study. *Trends in linguistics studies and monographs*, 35.

Oviatt, S., & Cohen, P. (2000). Perceptual user interfaces: multimodal interfaces that process what comes naturally. *Communications of the ACM, 43*, 45-53.

Padden, C., & Humphries, T. (1988). *Deaf in America: Voices from a Culture.* Cambridge: Harvard University Press.

Pandzic, I., & Forchheimer, R. (Eds.). (2003). *MPEG-4 facial animation: the standard, implementation and applications.* . Hoboken, NJ: John Wiley & Sons.

Schnepp, J. C., Wolfe, R. J., McDonald, J. C., & Toro, J. A. (2012). Combining emotion and facial nonmanual signals in synthesized american sign languag. *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility* (pp. 249-250). Dundee, UK: ACM.

The ASL Avatar Project Team at DePaul University. (2012, May 31). *Demonstration video of our ASL Avatar.* Retrieved December 24, 2017, from The American Sign Language Avatar Project: http://asl.cs.depaul.edu/demo.html

Theobald, B.-J., Bangham, J. A., Matthews, I., & Cawley, G. (2003). Evaluation of a talking head based on appearance models. *AVSP 2003-International Conference on Audio-Visual Speech Processing.*

Wolfe, R., Cook, P., McDonald, J., & Schnepp, J. (2009). *Toward a Better Understanding of Nonmanual Signals through Acquisition and Synthesis.* Frankfurt: Presented at Nonmanuals in Sign Languages (NISL).

Zorić, G., & Pandžić, I. (2005). A real-time lip sync system using a genetic algorithm for automatic neural network configuration. *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, (pp. 1366-1369).

Zorić, G., & Pandžić, I. (2006). Real-time language independent lip synchronization method using a genetic algorithm. *Signal processing, 86*(12), 3644-3656.